CHARACTERIZATION OF RIBONUCLEOTIDES EMBEDDED IN DNA

A Dissertation
Presented to
The Academic Faculty

by

Sathya Balachander

In Partial Fulfillment
of the Requirements for the Degree
Doctorate of Philosophy in the
School of Biological Sciences

Georgia Institute of Technology
December 2018

CHARACTERIZATION OF RIBONUCLEOTIDES EMBEDDED IN DNA

Approved by:

Dr. Francesca Storici, Advisor
School of Biological Sciences
Georgia Institute of Technology

Dr. Yuhong Fan
School of Biological Sciences
Georgia Institute of Technology

Dr. Yury Chernoff
School of Biological Sciences
Georgia Institute of Technology

Dr. Philip Santangelo
Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Natasha Degtyareva
National Institute of Environmental Health Sciences
National Institutes of Health

Date Approved: October 23, 2018

For my family and friends

ACKNOWLEDGEMENTS

I am extremely fortunate to have been a part of the Storici lab. I very thankful for having Dr. Francesca Storici as my PhD advisor. No matter how asinine the research idea or my question was, Francesca would sit with me, sometimes for hours, and steer me in the right direction. In addition to building my research profile, I had a great opportunity to attend international and national conferences, mentor students, and take part in all the other 'non-lab-related' activities. I spent more time with my lab mates than my family or friends, and I truly could not have survived graduate school without their support.

I wish to thank my thesis committee, Dr. Yury Chernoff, Dr. Natasha Degtyareva, Dr. Yuhong Fan and Dr. Phil Santangelo for their insight. Additionally, I am thankful to have been part of the School of Biological Sciences and for receiving their constant support. A huge thanks to Dr. Joe Lachance for letting me borrow his laptop and I wouldn't be typing this if not for his timely help. A very special thanks to Kyung Duk Koh, Taehwan Yang, Alli Gombolay, and Penghao Xu for being incredible and helpful, especially during the last couple of years.

I am extremely grateful and lucky to have some of the most loyal, genuine, quirky and lovable friends. Some of whom I spent all my days, including weekends I must say, with (Storici and Chernoff lab members), some I try to regularly meet (fellow GaTech and ATL folks), and all my friends in different time zones who are there for me no matter how silly and/or stressful the situation I am in.

Lastly, I am forever grateful for the unrelenting support from my parents, and my brother. Their unwavering support, optimism, and confidence in me are something that I do not take for granted. This PhD is a proof of all the collective sacrifices we have made over the years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AGS	Aicardi-Goutieres Syndrome
AP	Apurinic/apyrimidinic
APE1	Apurinic/apyrimidinic endonuclease I
AtRNL	*Arabidopsis thaliana* tRNA ligase
BER	Base excision repair
bp	Base-pair
CTD	C-terminal domain
dAMP	Deoxyadenosine monophosphate
dATP	Deoxyguanosine monophosphate
dCMP	Deoxycytidine monophosphate
dCTP	Deoxcytidine triphosphate
dF	Tetrahydrofuran abasic deoxyribonucleotide
dGMP	Deoxyguanosine monophosphate
dGTP	Deoxyguanosine triphosphate
DNA	Deoxyribonucleic acid
dNMP	Deoxyribonucleoside monophosphate
dNTP	Deoxyribonucleoside triphosphate
dRP	5’deoxyribose phosphate
ds	Double-strand
DSB	Double-strand break
dTMP	Deoxythymidine monophosphate
dTTP	Deoxythymidine triphosphate

dUMP	Deoxyuridine monophosphate
EDTA	Ethylenediaminetetracetic acid
EMSA	Electrophoretic mobility shift assay
F	Tetrahydrofuran
FBS	Fetal bovine serum
FEN1	Flap endonuclease 1
fmol	Femtomole
h	Hours
K	Kelvin
kb	Kilobases
kd	Knockdown
m	Meters
M	Molar
MEF	Mouse embryonic fibroblast
Mg	Magnesium
mg	Milligrams
min	Minutes
mL	Milliliters
mM	Millimolar
MMR	Mismatch repair
mtDNA	Mitochondrial DNA
NCE	Nuclear cell extracts
nDNA	Nuclear DNA
NER	Nucleotide excision repair
ng	Nanograms

NIR	Nucleotide incision repair
nM	Nanomolar
nmol	Nanomole
nt	Nucleotide
NTD	N-terminal domain
OGG1	8-Oxoguanine glycosylase
OH	Hydroxyl
oligo	Oligonucleotide
O/N	Overnight
P	Product
PAGE	Polyacrylamide gel electrophoresis
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
Pfu	*Pyrococcus furiosus*
PIP	PCNA-interacting protein
pM	Picomolar
pmol	Picomoles
Pol	Polymerase
r8oxoG	Ribose 8-oxo-guanosine
rAMP	Adenosine monophosphate
rAP	Ribo apurinic/apyrimidinic
rATP	Guanosine triphosphate
rCMP	Cytidine monophosphate
rCTP	Cytidine triphosphate
rDNA	Ribosomal DNA

RED	Ribonucleotide excision defective
RER	Ribonucleotide excision repair
rF	Tetrahydrofuran abasic ribonucleotide
rGMP	Guanosine monophosphate
rGTP	Guanosine triphosphate
RNA	Ribonucleic acid
RNase H	Ribonuclease H
rNMP	Ribonucleoside monophosphate
rNTP	Ribonucleoside triphosphate
ROS	Reactive oxygen species
rRNA	Ribosomal RNA
rUMP	Uridine monophosphate
S	Substrate
s	Second
siRNA	Short interfering RNA
SLE	Systemic lupus erythematosus
ss	Single-strand
ssDNA	Single-strand DNA
ssRNA	Single-strand RNA
Taq	*Themrus aquaticus*
Top1	Topoisomerase I
Ty	Transposable element
U	Unit
UMI	Unique molecular identifier
UNG1	Uracil-N-Glycosylase

UV	Ultraviolet
WCE	Whole cell extracts
w/v	Weight-to-volume
WT	Wild-type
YPD	Yeast extract/peptone/dextrose
α	Alpha
β	Beta
°	Degrees
° C	Degree Celsius
Δ	Deletion
δ	Delta
ε	Epsilon
γ	Gamma
η	Eta
μ	Mu
μL	Microliters
μm	Micrometers
μM	Micromolar

SUMMARY

Ribonucleotides (rNMPs) are the most abundant non-standard nucleotides in genomic DNA. Presence of rNMPs embedded in DNA alters the DNA structure, function and their properties, which ultimately may lead to genomic instability in the form of mutagenesis, replication stress and DNA breaks. Despite abundant evidence of the negative impact of rNMPs in DNA, not much is known about location and identity of rNMPs incorporated in genomic DNA. Here, our aims are to study the genome-wide distribution of rNMPs, and to characterize DNA repair mechanisms responsible for removal of rNMPs and modified rNMPs. To better understand the profile of rNMPs in DNA, we modified our current method, ribose-seq, to generate a robust and effective technique to capture rNMPs incorporated in DNA. Using our modified ribose-seq, in addition to mapping rNMPs in ribonuclease (RNase) H2 null cells at a much higher efficiency, we also determined the rNMP incorporation pattern from wild type DNA of budding yeast *S. cerevisiae* and *S. paradoxus*, and fission yeast *S. pombe.* Additionally, we explored the role of RNase H2 in cleaving modified rNMPs, such as abasic rNMPs. To study if RNase H2 can cleave at abasic rNMPs in DNA, we investigated whether eukaryotic RNase H2 is capable of recognizing abasic rNMPs. Also, we investigated the role of base excision repair (BER) enzymes in cleaving rNMPs and abasic rNMPs. We identity the role of apurinic/apyrimidinic endonuclease 1 (APE1) in cleaving abasic rNMPs, thus revealing a novel function of the BER pathway.

CHAPTER 1. INTRODUCTION

1.1 DNA and RNA structure

Nucleic acids are the building blocks of life. Nucleic acids differ in their structure of the sugar in their nucleotides. They are composed of nucleotides that consists of a 5-carbon sugar, a phosphate group, and a nitrogenous base. The 5-carbon sugar can be either deoxyribose or ribose, thus giving rise to deoxyribonucleoside 5'-monophosphates (dNMPs) or ribonucleoside 5'-monophosphates (rNMPs). dNMPs do not have a 2'-hydroxyl (OH) whereas rNMPs contains a 2'-OH group, thereby making it less stable (**Figure 1.1**). The base, adenine (A), cytosine (C) and guanine (G) are found in RNA and DNA, while thymine (T) occurs in DNA and uracil (U) occurs in RNA. Presence of a 2'-hydroxyl (OH) on ribose sugar of RNA makes it susceptible to strand breaks [2].

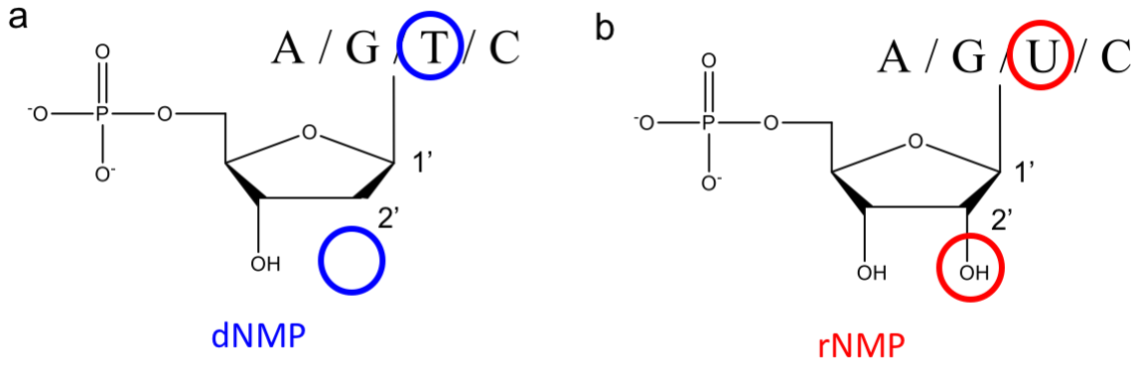


Figure 1.1 Deoxyribonucleotide vs Ribonucleotide. (a) Deoxyribonucleotide (dNMP) lacks a 2'-hydroxyl (OH) and has adenine (A), guanine (G), thymine (T) or cytosine (C) as one of its bases **(b)** Structure of ribonucleotide containing a 2'-OH, and having uracil (U) as one of the four bases, in place of thymine in dNMP.

1.2 Genome stability

DNA and RNA are two fundamental molecules of life. DNA is the genetic material that stores biological information. It is responsible for all of the basic cellular functions. DNA is transcribed into RNA. RNA has many regulatory functions in the cells; moreover, RNA is translated into proteins, and proteins help in maintaining major cellular functions. It is crucial to maintain the integrity and stability of the genome, as our DNA is constantly threatened by various exogenous agents like chemicals and UV radiation, and by endogenous agents such as DNA replication errors, dysfunctional DNA repair pathways, and oxidative stresses [3]. These agents can modify the chemical structure of DNA causing DNA lesions such as alkylation, oxidation, single and double-strand breaks (DSBs), or other types of modified or abnormal nucleotides [4, 5]. Among the different types of DNA lesions, rNMPs are the most commonly found non-standard nucleotides in DNA [6]. Genome stability is maintained by specialized DNA repair pathways that can remove the lesions and help reduce their effects. For example, DNA damage induced by UV result in bulky DNA adducts that are recognized by nucleotide excision repair (NER) [7], while specific non-bulky lesions in DNA are corrected by base excision repair (BER) [8]. Similarly, mismatch repair (MMR) helps in repairing faulty insertion, deletion and misincorporation of bases in DNA [9]. Improper repair of DNA lesions can lead to deleterious mutations, cell death, and diseases such as cancer [10, 11].

1.3 Presence of ribonucleotides in DNA

Studies have shown that the cellular concentrations of ribonucleoside 5’-triphosphate (rNTPs) are much higher than deoxyribonucleoside 5’-phosphate (dNTPs). In

Saccharomyces cerevisiae and human cells, the physiological concentrations of rNTPs are 36- to 190- fold molar excess than their corresponding dNTPs [12, 13]. As shown in **Figure 1.1**, rNMPs have an extra 2'OH as compared to dNMPs. If rNMPs are (mis)incorporated in DNA, the highly reactive 2'OH can attack the sugar-phosphate backbone causing internal threat to DNA, thus making it unstable [2]. Some DNA polymerases use a steric gate residue to discriminate against the 2'-OH in rNMPs to selectively insert dNMPs in DNA [14]. However, this process is not very efficient as DNA polymerases of bacteria, yeast and human can occasionally insert rNMPs in the DNA. *E. coli* DNA polymerases I, II, V, and the polymerase component of bacterial non-homologous end joining ligases can incorporate rNMPs [15-17]. rNMP incorporation by yeast replicative polymerases, α, δ, and ε in DNA happens with an insertion ratio (rNMP/dNMP) of 1/625, 1/5,000 and 1/1,250, respectively [12]. On measuring the physiological *in vitro* nucleoside triphosphate concentrations, these polymerases can introduce more than 13,000 rNMPs in yeast genome during each round of replication [12]. *In-vitro* studies showed that human Pol δ can misincorporate 1 rNMP for every 2,000 dNMPs suggesting that human Pol δ may introduce more than a million rNMPs into nuclear genome per replication cycle [18]. Human Pol η and Pol ι also were shown to insert rNMPs into DNA [18-20].

Additional sources of rNMP incorporation include incomplete Okazaki fragment maturation during replication, and oxidative damage. RNA primer initiates the formation of Okazaki fragments during lagging strand synthesis [21]. This RNA fragment is later removed in the process of Okazaki fragment maturation, but incomplete removal may lead to rNMPs to be incorporated in DNA [22]. Oxidation of the deoxyribose sugar into ribose

can also lead to formation of rNMPs. It has been shown, *in vitro* and *in vivo*, that reactive oxygen species (ROS) can cause multiple modifications to dNMPs, including oxidation of the deoxyribose sugar forming ribose, to result in formation of rNMPs and other modified nucleotides [23]. Treatment of DNA under Fenton reaction conditions can result in conversion of dNMPs to rNMPs [24].

Considering the abundance of rNMPs, there is a good likelihood of having modified rNMPs, such as abasic rNMPs also present in DNA. Abasic sites are formed by spontaneous hydrolysis of the *N*-glycosidic bond in DNA or following removal of a damaged base [8]. As mentioned earlier, rNMPs were shown to form during oxidative DNA damage both *in vitro* and *in vivo* [24]. Therefore, it is possible that abasic and oxidized dNMPs can be converted to abasic rNMPs. Spontaneous depurination is estimated to occur 1,000 times slower in RNA than DNA [25]. The rate of depurination in DNA under physiological conditions is estimated to be 1 out of 100,000 purines every cell cycle, which gives us a rate of 10,000 abasic sites per day in human cells [26]. Due to the abundance of rNMPs in the genome, with more than 600,000 in yeast and more than 100 million in human DNA [27], hence the possibility of abasic rNMPs present in DNA is very realistic.

1.4 Effects of ribonucleotides

Unrepaired rNMPs can have deleterious effects in DNA. Due to the highly reactive 2'-OH group in the sugar, presence of rNMPs in DNA sensitizes the backbone to DNA breaks [2]. rNMPs incorporated in DNA can become template for the next round of replication, where DNA polymerases α, δ or ε can bypass these template containing rNMPs [18, 28]. Short

tracts of rNMPs in DNA can serve as templates for DNA synthesis and can propagate mutation upon replication in *E. coli,* budding yeast, and human embryonic kidney cells [29-31]. The presence of rNMPs in DNA can stall replication fork progression, thereby leading to replication stress [20, 32]. Also, misincorporated rNMPs in yeast and human DNA, if left unrepaired can impede cell cycle progression and result in checkpoint activation [33-35]. Structurally, the presence of rNMPs can alter the helical parameters and result in a transition from B-form to A-from [36]. Misincorporated rNMPs can affect the elastic properties of duplex DNA and possibly alter the way DNA binding proteins interact with DNA [37]. Additionally, presence of rNMPs may disrupt the nucleosome binding to DNA *in vitro,* and also affect the proper assembly of nucleosomes [38, 39].

Although there is abundance of evidence suggesting the detrimental effect of rNMPs, there are a few known positive physiological roles of rNMPs. Presence of rNMPs in DNA helps signal the activation of mismatch repair (MMR). Presence of rNMPs during replication marks the nascent strands, thus acting as strand discrimination signals in eukaryotes [40, 41]. Presence of a two-rNMP imprint in the mating type locus of *Schizossacharomyces pombe* drives recombination process that results in mating type switching between minus (M) and plus (P) type [42]. In order to process the rNMPs there are multiple repair pathways that can cleave and repair the misincorporated nucleotides.

1.5 Repair mechanisms of rNMPs in DNA

As rNMPs incorporated in DNA can have adverse effects on the genome stability and integrity, there are several repair mechanisms that can target and remove the rNMPs in genomic DNA.

1.5.1 Ribonucleotide excision repair (RER pathway)

Ribonucleotides are efficiently removed by the ribonucleotide excision repair (RER) pathway. The process is initiated by ribonucleases (RNases) H, which are the primary enzymes that can cut rNMPs in DNA and also cleave the RNA strand of RNA/DNA hybrids [43]. There are two types of RNase H: type I/1 and type II/2. RNase H1 is active on stretches of four or more consecutive rNMPs embedded in DNA [44]. On the other hand, RNase H2 can remove a single rNMP and stretches of rNMPs in DNA, as well target RNA-DNA hybrids, as shown in **Figure 1.2** [44, 45]. In prokaryotes and eukaryotes, RNase H I/1 has only one subunit. Prokaryotic RNase HII has one subunit whereas eukaryotic RNase H2 has three different subunits, the catalytic subunit (Rnh201 in yeast, and RNase H2A in mammals) and two additional subunits (Rnh202 and Rnh203 in yeast, and RNase H2B and RNase H2C in mammals) that are needed for catalysis.

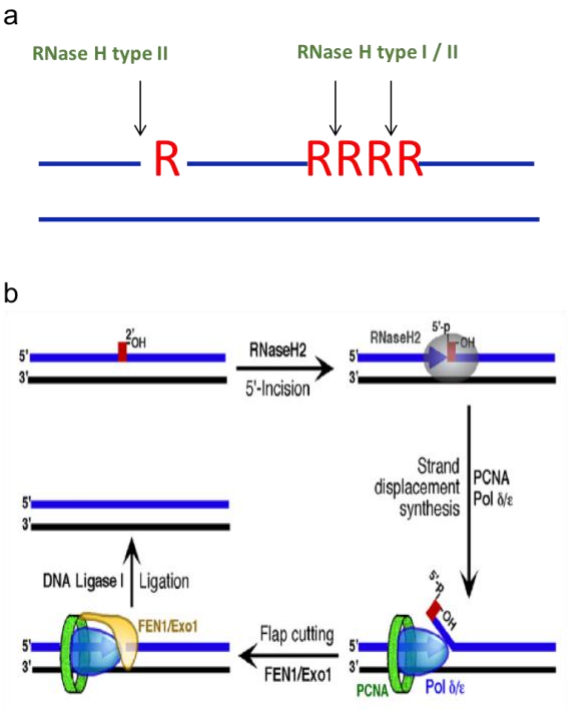


Figure 1.2 RNase H2 removal of rNMPs in ribonucleotide repair pathway (RER). (**a**) RNase H type I can recognize four or more stretches of rNMPs, whereas RNase H type II can recognize single and multiple stretches of rNMPs in DNA. (**b**) RER pathway is initiated by incision to the 5' end of rNMP by RNase H2, which is followed by strand displacement synthesis by DNA polymerase, nucleolytic flap processing, and ligation. rNMPs are shown here in red. Modified from [1].

The process of RER is initiated by RNase H2-mediated recognition of rNMPs. RNase H2 cleaves the DNA 5' of the rNMP, resulting in a DNA containing 3'-hydroxy and 5'-phosphoribonucleotide ends. This is followed by strand displacement synthesis of the nicked DNA strand by DNA Pol δ or Pol ε, followed by flap cleavage by flap endonuclease (FEN) 1, and ligation of the broken ends by DNA ligase 1 [1]. One of the subunits of RNase H2, Rnh202 contains a PCNA-interacting motif (PIP) box which binds to proliferating cell nuclear antigen (PCNA), a complex that plays an essential part at the replication fork through recruitment of many enzymes, so as to localize RNase H2 to the DNA replication

fork [46, 47]. The presence of the PIP motif in RNase H2 may suggest the coupling of RER with DNA replication. In human cells, RNase H2 is shown to be recruited to DNA replication and repair foci in both a PCNA-dependent and PCNA-independent manner [48].

Having a compromised RER pathway can result in deleterious effects. Deletion of genes encoding RNase H2 subunits in *S. cerevisiae* have shown to cause an increased rate of spontaneous recombination [49, 50]. Also, gross chromosomal rearrangements and gene conversion events are elevated in RNase H2-deficient yeast strains [49-52]. Null mutations in RNase H1 or RNase H2 can be tolerated in yeast, but deletion of either RNase H enzymes in mice is embryonic lethal and activate a p53-dependent damage response [44]. Murine RNase H2-null embryonic fibroblasts accumulate over one million rNMPs in their genome [53]. In humans, mutations in any of the three subunits of RNase H2 are associated with neurodegenerative disorder, such as Aicardi-Goutieres Syndrome (AGS), and autoimmune disorder such as systemic lupus erythematosus (SLE) [54, 55].

1.5.2 Other DNA repair mechanisms involved in rNMP removal

RNase H2 is the major repair enzyme to cleave rNMPs incorporated in DNA, but in the absence or failure of RER, rNMPs can be removed by DNA topoisomerase I (TopI) through topoisomerase I-mediated rNMP removal [35, 50, 56, 57]. Top1 is critical for resolving DNA supercoils that accumulate during replication and transcription, and also functions in removing rNMPs in DNA [58]. Top1 cleaves at rNMPs generating a single-strand DNA break containing DNA ends with 5'-OH and 2'-3'-cyclic phosphate ends. These ends cannot be processed by ligases but can be reversed by Top1 [57]. When this nicking

happens in DNA repeated sequences, a second Top1-dependent incision occurs, which can lead to loss of one or more parts of the repeated sequences [59].

Additionally, there are other DNA repair pathways that can help in removing rNMPs, although on specific substrates. The mismatch repair (MMR) system can recognize rNMP/dNMP mismatches and small insertions/deletions in *E. coli* and budding yeast [29]. The nucleotide excision repair (NER) mechanism can recognize a vast array of backbone distortions in DNA [60]. Rad14 of the NER pathway can target an rNMP in DNA in budding yeast (our lab unpublished data). Unlike in yeast, NER in human cells does not target rNMPs [61]. A minor fraction of the misincorporated rNMPs can be removed by the 3'-5' exonucleolytic function of DNA polymerases. Yeast and human Pol ε have a weak exonuclease activity on rNMPs [62, 63]. Also, yeast and human Pol δ exhibit a very little rNMP proofreading ability [18].

1.5.3 Base excision repair (BER) pathway

Base excision repair (BER) is a major pathway for removing damaged bases in DNA. It is known to remove a wide spectrum of oxidative lesions and abasic dNMPs in genomic DNA [64, 65]. BER is initiated by DNA glycosylases. DNA glycosylase flips the damaged base out of the double helix and cleaves the *N*-glycosylic bond of the damaged base, thereby leaving an apurinic/apyrimidinic (AP) site, also called as abasic site. These AP sites are recognized by AP endonuclease (APE) I, which cleaves the 5'-end of the AP site to yield a 3'hydroxyl adjacent to a 5'deoxyribosephosphate (dRP). Following the removal of AP site, the break is processed by either a short-patch or a long-patch, by Pol β, FEN1 and DNA ligase 1. Recent studies have shown that human APE1 can endonucleolytically cleave

abasic single-stranded RNA [66, 67]. The role of BER in targeting rNMPs or abasic rNMPs in DNA remains to be investigated.

1.6 rNMP mapping techniques

To map the location of rNMPs in the genome, four methods have been developed. After initial DNA fragmentation, embedded ribose-sequencing (emRibo-seq) uses RNase H2 to cleave 5' of an rNMP *in vitro*, generating a 3'-OH which can be tagged, thereby sequencing the downstream nucleotides of an rNMP. Three other techniques, hydrolytic end-sequencing (HydEN-seq), polymerase usage technique (PU-seq) and ribose-seq, our method, after the initial DNA fragmentation, use alkaline hydrolysis treatment to cleave the 3'end of rNMP, and resulting in 2'3'-cyclic phosphate and 5'-OH ends [68-71]. HydEN-seq and PU-seq tags the 5'-OH end fragment, thereby capturing the dNMP downstream from the rNMP. Ribose-seq uses *Arabidopsis thaliana* tRNA ligase (AtRNL) to ligate the 2'3'-cyclic phosphate end of an rNMP to the 5'-phosphate of the same fragment, thus directly capturing the rNMP and not the upstream or downstream dNMP of rNMP sites. Additionally, ribose-seq does not capture unprocessed Okazaki fragments, RNA primers, nicks or abasic sites in DNA.

All four techniques were developed using RNase H2-null strains of yeast, either *S. cerevisiae* or *S. pombe*, as there is an increase in rNMPs within DNA in RNase H2 defective cells. Despite having these methods to map rNMPs, there is a need to obtain a wide spectrum of rNMPs incorporation in genomic DNA of cells having not only inactive but also active RER system.

1.7 Research goals

To better understand the biological function and consequence of rNMPs in DNA, we need to know the identity and location of the rNMPs that are incorporated in genomic DNA. Additionally, to investigate if there are DNA repair pathways that can target modified rNMPs in DNA, we set up the following research goals.

1.7.1 Analysis of rNMPs embedded in genomic DNA

To develop a more robust system for studying the location, sequence context, and hotspots of rNMPs embedded in DNA, we set out to optimize the ribose-seq technique. Using this modified technique, we can get a better understanding of rNMP incorporation in wild type and RNase H2-null strains of *S. cerevisiae.* Additionally, we can also analyse and compare the profile of rNMP spectra of WT and several RNase H2 defective cells of *S. cerevisiae, S. paradoxus* and *S. pombe.*

1.7.2 To study the role of RER and/or BER in recognizing and cleaving rNMPs and/or abasic rNMPs.

Although RNase H2 can effectively target and remove rNMPs that are embedded in DNA, there is no information whether RNase H2 can cleave modified rNMPs, such as abasic rNMPs. Here we study if yeast, mouse and human RNase H2 can target abasic rNMPs. Additionally, we investigate if UNG1 and APE1 of the BER pathway can recognize rNMPs or abasic rNMPs.

CHAPTER 2. RIBOSE-SEQ – GLOBAL MAPPING OF RIBONUCLEOTIDES EMBEDDED IN DNA

The study in Chapter 2 consists of the work published in *Nature Methods* 12 (2015) 251-257. In addition to assisting in the overall development of ribose-seq, the assay and analysis to determine the role of UNG1 in cleaving rUMP in DNA were my contributions to this study.

Koh, K.D.[1†], Balachander, S.[1], Hesselberth, J.R.[2,] & Storici F.[1]

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA.

[2]Department of Biochemistry and Molecular Genetics, University of Colorado, Anschutz, Medical School, Aurora, CO 80045, USA.

[†]Present address: School of Medicine, University of California, San Francisco, CA 94143, USA.

2.1 Abstract

Abundant ribonucleotide incorporation during DNA replication, repair, and damage has profound consequences for genome stability, but the global distribution of ribonucleotide incorporation is unknown. We developed a method, Ribose-seq, to capture DNA fragments with terminal 2′,3′-cyclic phosphates or 2′-phosphates, which are unique products generated by alkaline cleavage of DNA at embedded ribonucleotides. High-throughput sequencing of these captured fragments from yeast *Saccharomyces cerevisiae* DNA revealed widespread and biased ribonucleotide distribution, with a strong preference for cytidine and guanosine, and identified hotspots of ribonucleotide incorporation in nuclear and mitochondrial DNA. Ribonucleotides were primarily incorporated on the leading strand of nuclear DNA and were present upstream of G+C-rich tracts in the mitochondrial genome. Ribose-seq establishes a powerful tool for systematic profiling of ribonucleotide incorporation in genomic DNA.

2.2 Introduction

Genomic DNA contains embedded ribonucleotides (rNMPs) that are incorporated during DNA replication, repair, and damage (reviewed in [6]). rNMPs were initially found at specific DNA loci in mouse and human mitochondrial DNA [72] and the mating type locus of fission yeast [73], but have since been detected in a variety of cell types [74]. Many DNA polymerases can incorporate rNMPs into DNA, including the human replicative DNA polymerase (Pol) δ [18] and mitochondrial Pol γ [75], budding yeast nuclear replicative Pol α, δ, and ε [12], *Escherichia coli* polymerase V [16], and the polymerase component of bacterial non-homologous end joining ligases [17]. rNMP incorporation

could also be a consequence of incomplete maturation of Okazaki fragments during lagging strand synthesis in DNA replication [22]. Moreover, generation of hydroxyl radicals ′OH during oxidative stress can modify DNA deoxyribose sugars to ribose, forming rNMPs in DNA both *in vitro* and *in vivo* [24].

Ribonuclease (RNase) H enzymes cleave rNMPs incorporated into DNA. RNase H type 1 (H1/HI) recognizes long stretches (≥4 nt) of rNMPs in RNA/DNA hybrids, whereas RNase H type 2 (H2/HII) cleaves even singly incorporated rNMPs in DNA [44]. Both RNases H1 and H2 are present in the nucleus, and only RNase H1 has been found in mitochondria [44, 75]. RNase H2/HII initiates ribonucleotide excision repair (RER), the major rNMP repair mechanism in bacterial DNA and in eukaryotic nuclear DNA [1, 44]. Inactivation of RNase H2/HII enabled detection of unexpectedly high amounts of rNMPs in genomic DNA. Quantitative measurements of rNMPs in DNA from RNase H2-deficient mouse embryonic fibroblasts (MEF) identified >1 million rNMPs per genome, suggesting that rNMPs are the most common non-canonical nucleotides in dividing mouse cells [53]. Similar measurements conducted for genomic DNA derived from RNase H2-deficient (*rnh201*Δ) budding yeast estimated a few thousand rNMPs incorporated per genome per cell cycle [35, 41], and RNase HII-null *Bacillus subtilis* cells have high levels of incorporated rNMPs [76]. Embedded rNMPs in DNA have highly reactive 2′-hydroxyl groups, altering its properties, structure, and function [2, 37] and leading to genome instability [35, 77-79]. Also, in humans, mutations in any of the three subunits of RNase H2 are associated with the neurological syndrome Aicardi-Goutieres (AGS) [54].

Despite abundant evidence for the frequent incorporation of rNMPs in DNA, a detailed picture of how rNMPs are incorporated throughout a genome is lacking. Here, we introduce Ribose-seq: a technique for mapping rNMPs in genomic DNA.

2.3 Materials and Methods

2.3.1 Construction of yeast strains

Yeast strains used in this study are presented in **Table A.1**. Isogenic yeast haploid strains KK-100, KK-174, KK-107, and KK-120 were derived from E134 (*MATα ade5-1 lys2-14A trp1-289 his7-2 leu2-3,112 ura3-52*) [80]. KK-100 was made from E134 by deletion and replacement of *RNH201* via transformation with a PCR product containing the *hygMX4* cassette flanked by 50 nucleotides of sequence homologous to regions upstream and downstream of the *RNH201* ORF. KK-174 was constructed from KK-100 by deletion and replacement of *RNH1* via transformation with a PCR product containing the *kanMX4*. KK-107 was generated by introducing *pol2-4* mutation into KK-100 via integration-excision using plasmid YIpJB1 [81]. KK-120 was made by introducing *pol3-5DV* mutation into KK-100 via integration-excision using plasmid p170-5DV [82].

Isogenic yeast haploid strains KK-30, KK-125, KK-164, and KK-170 were derived from FRO-767,768 (*ho*Δ *hml*Δ::*ADE1 MATa-inc hmr*Δ::*ADE1 ade1 leu2-3,112 lys5 trp1*::*hisG ura3-52 ade3*::*GAL*::*HO leu2*::*HOcs mata*Δ::*hisG*) [31]. KK-30 was made from FRO-768 by reversion of *ade3*::*GAL*::*HO* to intact *ADE3* via transformation with a PCR product containing *ADE3*, followed by replacement of *RNH201* with the *hygMX4*. KK-125 was constructed from KK-30 by replacement of *RNH1* with the *kanMX4* cassette. KK-164 was generated from KK-125 by replacement of *UNG1* with the *natMX4* cassette. KK-170 was

made by introducing *pol2-M644G* mutation into KK-30 via integration-excision using plasmid p173-M644G [83].

Isogenic yeast haploid strains KK-158,159 were derived from FRO-767,768 [31]. KK-158,159 were constructed from FRO-767-768 by replacement of *UNG1* with the *hygMX4* cassette.

2.3.2 AtRNL ligation assay

Ribonucleotide (rNMP)-containing DNA oligonucleotide (oligo), Lig.47.R (see **Table A.2**), and its DNA-only control, Lig.47.D, were 5′ end-labeled with [γ-^{32}P]ATP (PerkinElmer) by T4 polynucleotide kinase (New England BioLabs). Alkali treatment was carried out in 0.3 M NaOH for 2 hr at 55 °C. The resulting solution was neutralized and diluted. 100 nM of alkali-treated 5'-radiolabeled products were incubated in 50 mM Tris-HCl, pH 7.5, 40 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 30 µM ATP (Sigma-Aldrich), and 1 µM AtRNL [84] for 1 hr at 30 °C. After dilution, the ligated products and remaining substrates were treated with T5 exonuclease (NEB) for 2 hr at 37 °C. Aliquots were withdrawn after appropriate steps and quenched with 90% formamide. The products were analyzed by 15% (w/v) polyacrylamide, 8 M urea gel electrophoresis (urea-PAGE). 20-100 Oligo Length Standard (Integrated Device Technology) was used as a ladder. Following electrophoresis, gels were exposed to a phosphor screen overnight. Images were taken with Typhoon Trio$^+$ (GE Healthcare) and obtained with ImageQuant (GE Healthcare). Band intensities were quantified by Multi Gauge V3.0 (Fujifilm).

2.3.3 3’ base bias for AtRNL ligation assay

rAMP, rGMP, rUMP, and rCMP-containing DNA oligos (Lig.30.rA, Lig.30.rG, Lig.30.rU, and Lig.30.rC, respectively; see **Table A.2**) were 5′ end-labeled with either hot [γ-^{32}P]ATP (PerkinElmer) or cold ATP (Sigma-Aldrich) by T4 polynucleotide kinase (NEB). Each of hot rNMP-containing 5′-radiolabeled DNA oligos was mixed with other three cold equimolar DNA oligos. The mixtures were treated with 0.3M NaOH for 2 hr at 55 °C, neutralized, and diluted. 100 nM of alkali-cleaved products (25 nM of each base) were then incubated in 50 mM Tris-HCl, pH 7.5, 40 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 30 μM ATP (Sigma-Aldrich), and either 1 μM or 200 nM AtRNL [84] for 1 hr at 30 °C. After dilution, the resulting products were treated with T5 exonuclease for 2 hr at 37 °C. Aliquots were withdrawn after appropriate steps, quenched, and analyzed by urea-PAGE.

2.3.4 rNMP bypass assay

A DNA primer oligo, ByPrim (see **Table A.2**), was 5′ end-labeled with [γ-^{32}P]ATP (PerkinElmer) by T4 polynucleotide kinase (NEB). The 5’-radiolabeled primer was annealed to either rCMP- or rUMP-containing template oligo (ByTemp.rC or ByTemp.rU, respectively). 100 nM of annealed substrates were incubated in 1X HF Buffer, 2 mM dNTPs, and 0.2 units of Phusion High-Fidelity DNA Polymerase (NEB) for 30 sec at 72 °C. The reactions were quenched and analyzed by urea-PAGE. Bypass probability was calculated as the band intensity at the +1 position plus all longer products divided by the intensity at the −1 position (preceding the rNMP) plus all longer products, as described in Kokoska *et al.*, 2003 [85].

2.3.5 Double-stranded break repair (DSB) assay with rNMP-containing oligos

Transformation with rNMP-containing DNA oligos, LEU2.rG and LEU2. rU (see **Table A.2**) and DNA-only oligos, LEU2.D and LEU2.dU, were done as described in Storici *et al*., 2007 [31]. Cells from each oligo transformation were plated to selective Leu^- medium. For each transformation, 20 Leu^+ transformants were selected. Colony PCR was performed on those transformants, amplifying with primers, LEU2.3 and LEU2.6, a 900-bp region in *LEU2* locus, where a new StuI restriction site is expected. The resulting PCR products were treated with StuI (NEB) and analyzed by agarose gel electrophoresis to check the presence of the StuI restriction site.

2.3.6 Ribose-seq library construction to map rNMPs in DNA

Genomic DNA from *S. cerevisiae* cells grown in liquid rich medium containing yeast extract, peptone, and 2% (w/v) dextrose (YPD) for 2 days to stationary phase was extracted following the protocol "Preparation of Yeast Samples" in QIAGEN Genomic DNA Handbook. Genomic-tip 500/G (QIAGEN), Genomic DNA Buffer Set (QIAGEN), Proteinase K (QIAGEN), RNase A (QIAGEN), and Lyticase (Sigma-Aldrich) were used to extract genomic DNA from *S. cerevisiae* cells. Extracted genomic DNA was digested with SspI, DraI, and EcoRV (NEB) overnight at 37 °C to create a population of 500–3,000 bp genomic fragments with an average size of ~1.5 kb. Assuming that rNMPs, if present, could be located in any position of each genomic fragment, an average of 1.5 kb allows a reasonable window of rNMP capture. Following confirmation of digestion by Experion Automated Electrophoresis System (BIO-RAD), the fragments were tailed with dATP (Sigma-Aldrich) by Klenow fragment (exo^-) (NEB) for 30 min at 37 °C. The resulting

products were purified by spin-column (QIAGEN) and then ligated to preannealed double-stranded (ds) adaptors (Adaptor.L/Adaptor.S, see **Table A.2**) that contain single dT overhangs and a randomized 8-base unique molecular identifier (UMI) by T4 DNA ligase (NEB) overnight at 15 °C. The products were purified using AMPure XP beads (Beckman Coulter). All subsequent purifications were done using AMPure XP beads. The adaptor-ligated DNA fragments were incubated in 0.3 M NaOH for 2 hr at 55 ˚C to expose 2′,3′-cyclic phosphate and 2′-phosphate termini of DNA at rNMP sites, followed by neutralization and purification. The resulting single-stranded (ss) fragments were incubated in 50 mM Tris-HCl, pH 7.5, 40 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 30 µM ATP (Sigma-Aldrich), and 1 µM AtRNL [84] for 1 hr at 30 °C, followed by purification. The products and remaining fragments of DNA were treated with T5 exonuclease (NEB) for 2 hr at 37 °C to degrade the background of unligated, linear ss DNA, leaving self-ligated ss DNA circles intact. Treatment with the 1 µM Tpt1 [84] in 20 mM Tris-HCl, pH 7.5, 5 mM MgCl2, 0.1 mM DTT, 0.4% Triton X-100, and 10 mM NAD^+ (Sigma-Aldrich) for 1 hr at 30 °C was used to remove the 2′-phosphate remaining at the ligation junction. After purification and resuspension, the libraries were PCR-amplified with one of the bar-coded primers, PCR.1.Index1-4, and PCR.2 (see **Table A.2**) using either Phusion High-Fidelity DNA Polymerase (NEB) or EconoTaq DNA Polymerase (Lucigen), confirmed by 6% polyacrylamide gel electrophoresis (PAGE), purified, and pooled for analysis by Illumina sequencing. 100 bp DNA Ladder (NEB) was used as a ladder. SYBR Gold Nucleic Acid Gel Stain (Life Technologies) was used to stain PAGE gels for visualization under ultraviolet light. Ribose-seq does not capture RNA primers of Okazaki fragments because the 5'-most rNMP is a 5'-triphosphate [86], and T4 DNA ligase used to attach the

sequencing adaptors absolutely requires a 5'-monophosphate [87]. Moreover, the rest of the primers are reduced to single nucleotides upon alkali treatment, and they will have no adaptor sequence ligated on. Ribose-seq also does not detect rNMP positions derived from residual RNA molecules or RNA/DNA hybrids not embedded in DNA (like cDNA) nor DNA abasic sites, which could have been ligated to the adaptor sequence by T4 DNA ligase. Following alkali treatment, RNA stretches are reduced to single nucleotides which are removed in subsequent purification steps; even if the 5-most rNMP is captured, the rNMP-containing ss circle would not have any sequence to be aligned to the reference genome. Abasic sites undergo both β- and δ-eliminations to yield 5′-phosphate and 3′-phosphate ends [88, 89], which cannot be ligated by AtRNL. Because of the nature of alkaline hydrolysis within a stretch of rNMPs embedded in DNA, our ribose-seq captures only the 5'-most rNMP of the stretch of two or more rNMPs. Moreover, ribose-seq does not require rNMPs to be present at the same location from cell to cell, and can identify incorporated rNMPs with single-base precision.

2.3.7 DNA sequencing

Indexed sequencing libraries were mixed at equimolar concentrations and normalized to 10 nM. Libraries were sequenced on an Illumina MiSeq, and 50 cycle single-end reads were collected. Raw sequencing reads are available at NCBI GEO [90] under accession GSE61464.

2.3.8 Alignment of sequence and processing

Reads were aligned to the *S. cerevisiae* genome (sacCer2) with bowtie using two different settings to report uniquely aligning and multiple aligning reads (`-m 1` and `--all`,

respectively). Aligned reads in BAM format were processed to remove PCR duplicates using umitools (https://github.com/brwnj/umitools), which filters reads that contain duplicate UMIs and reports reads with unique UMIs. Reads in this study had an 8-base UMI incorporated during ligation, corresponding to the first 8 cycles of raw FASTQ sequence. Following UMI removal, read depths at each 5′ position were calculated with BEDTools [91, 92].

2.3.9 Nucleotide frequencies

Nucleotide frequencies for mapped rNMP positions (*i.e.* the 5′ position of each aligned read) were calculated and normalized to genome frequencies (nuclear and mitochondrial genomes in sacCer2). The identity of the rNMP base is the reverse complement of the 5′ base of each read. Nucleotide frequencies of downstream sequences of incorporated rNMPs, including +1 position, cannot be affected by our approach of capturing rNMPs in DNA because the rNMPs and their upstream sequences are captured, sequenced and aligned to the reference genome.

2.3.10 Replication corelations

The density and identity of rNMPs present on leading and lagging strands were calculated relative to annotated origins of replication [93] and were further categorized by replication timing [94]. Data were filtered for specified replication timings (e.g. 25 minutes following release into S phase) and distances relative to the middle of each ARS annotation (e.g. 5.0 kb up- and downstream of each ARS).

2.3.11 Determination of hotspots of rNMP incorporation in genomic DNA

Two different analyses were conducted to identify hotspots of rNMP incorporation in genomic DNA. Peak calling was performed with macs2 (version 2.1.0.20140616) [95] with specific parameters (--keep-dup all --nomodel -s 25 --extsize 5 --call-summits). Peaks of length greater than 1,000 were filtered from further analysis, and remaining peaks with a *q*-value less than 0.001 were selected. A second analysis involved finding positions of rNMPs within the locus of interest with ribose-seq signal greater than the mean plus three standard deviations for each library from *rnh201*Δ (KK-100), *rnh201*Δ (KK-100, EconoTaq), *rnh201*Δ (KK-30), *rnh1*Δ *rnh201*Δ (KK-174), and *rnh1*Δ *rnh201*Δ (KK-125) cells.

2.3.12 Data presentation and statistics

Graphs were made using GraphPad Prism 5 (GraphPad Software). Nonparametric two-tailed Mann-Whitney *U*-test [96] was implemented for statistical analysis of AtRNL ligation efficiencies, rNMP bypass probabilities, and the percentages of StuI-cut Leu^+ transformants in DSB repair assay. Chi-square goodness of fit test[96] was used for statistical comparison of the distribution of rNMP reads to the expected Poisson distribution.

2.4 Results

2.4.1 Ribose-seq: Capturing rNMPs in DNA

Our ribose-seq approach captures rNMP-terminated single-stranded (ss) DNA fragments generated by alkaline cleavage of rNMPs in DNA (**Figure 2.1** and **Figure A.1**). We

exploited the unique ligation mechanism of *Arabidopsis thaliana* tRNA ligase (AtRNL), normally involved in tRNA maturation. AtRNL converts 2′,3′-cyclic phosphate ends of RNA to 2′-phosphate and ligate these to 5′-phosphate ends of RNA [84, 97] or DNA [97]. We demonstrated that AtRNL captures 2′,3′-cyclic phosphate or 2′-phosphate terminus of DNA derived from alkaline cleavage of a DNA oligonucleotide (oligo) at an embedded rNMP, ligating the 2′-phosphate end to the 5′-phosphate terminus of the same DNA molecule and producing a ss DNA circle containing an embedded rNMP. Self-ligation was strongly preferred over dimerization, as linear dimers were not detected (**Figure 2.1a**). Further, these ss DNA circles are resistant to T5 exonuclease enabling their enrichment relative to unligated linear DNA upon exonuclease treatment (**Figure 2.1a**). We did not observe any bias for the 3′ rNMP substrate of AtRNL, which captured an embedded rAMP, rCMP, rGMP, or rUMP with equal efficiency (P values > 0.05 in each case) (**Figure A.2** and **Table A.3**), nor was any bias observed in a previous study [98]. These data indicate that self-ligation is favored for AtRNL on 2′-phosphate-terminated ss DNA fragments as small as 22 nt (**Figure 2.1a** and **Figure A.2**), thus facilitating library construction and high-throughput DNA sequencing.

We applied ribose-seq to identify rNMPs embedded in nuclear and mitochondrial DNA of RNase H2-deficient budding yeast (strain KK-100, **Table A.1**) [6]. Genomic DNA was extracted from cells grown to stationary phase, and a mixture of three blunt-end restriction enzymes was used to fragment the DNA. Application of our rNMP-capture scheme (**Figure 2.1b**) yields a library of DNA molecules (**Figure A.3a**) with an average size of ~350 bp, each of which maps to a single site of rNMP incorporation and its upstream sequence. In control experiments, we found that exclusion of either AtRNL (**Figure A.3a**) or alkali

treatment (**Figure A.3b**) prevented library formation, validating that captured molecules derive from rNMPs embedded in DNA.

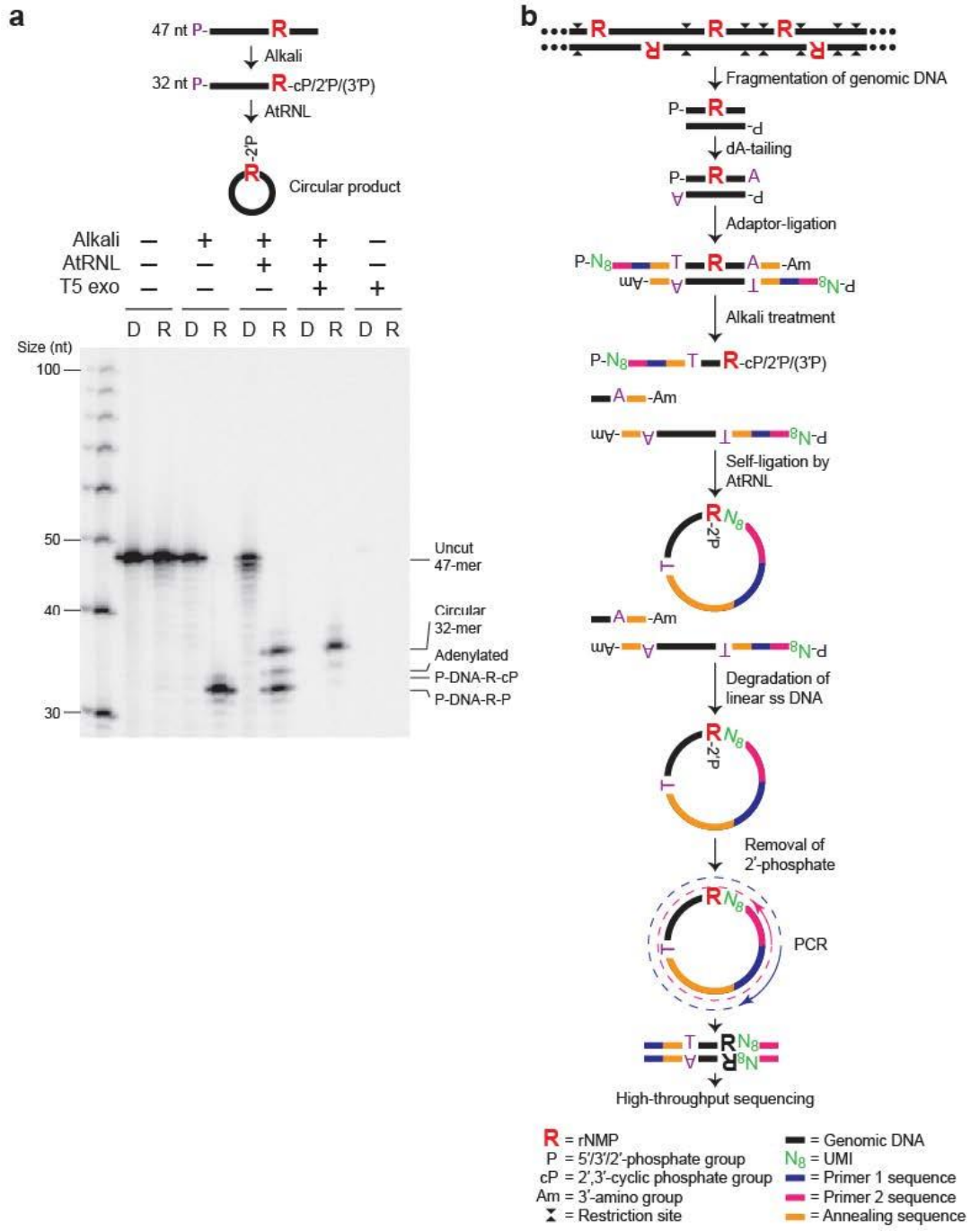


Figure 2.1 Capture of rNMPs in DNA via AtRNL for ribose-seq. (**a**) AtRNL was used to capture 2′,3′-cyclic phosphate or 2′-phosphate termini of DNA generated by alkaline cleavage of a single rGMP in the 32nd position of 5′-radiolabeled 47-nt ss DNA oligo, Lig.47.R (see **Table A.2**), 'R'. A DNA-only oligo, Lig.47.D, was used as a control, 'D'. 5′-radiolabel is indicated by 'P' in purple. Treatment with T5 exonuclease was performed to confirm the presence of circular ligation product while the unligated, linear substrates are degraded. First left lane, ss DNA ladder. Ligation efficiency was about 50%, as

expected, because there is a mixture of 2′-phosphate and 3′-phosphate ends upon alkaline cleavage at the rNMP in DNA. (**b**) Ribose-seq method for mapping rNMPs in genomic DNA. Genomic DNA is initially fragmented, dA-tailed, and ligated to an UMI-containing sequencing adaptor. Alkali treatment denatures the DNA and cleaves at rNMPs in DNA, exposing 2′,3′-cyclic phosphate and 2′-phosphate termini, which are self-ligated to 5′-phosphate ends by AtRNL. The linear, unligated fragments are degraded by T5 exonuclease while the remaining rNMP-captured, circular DNA molecules, upon removal of 2′-phosphate at the ligation junction by a 2′-phosphotransferase Tpt1, are PCR-amplified and sequenced.

2.4.2 Distribution and base composition of rNMPs in *S. cerevisiae* genome

Ribose-seq library prepared from *rnh201*Δ cells (KK-100) was sequenced to a depth of ~2 million reads, which were mapped to the yeast *S. cerevisiae* genome, allowing us to define rNMP locations along yeast nuclear and mitochondrial DNA with single-nucleotide resolution. This analysis uncovered widespread rNMP incorporation with a coverage of 0.449 and 19.5 rNMP reads/kb in the nuclear and mitochondrial genome, respectively (**Figure 2.2a** and **Table A.4**). While broadly scattered, the rNMP sites in the nuclear and mitochondrial DNA were not randomly distributed (**Figure 2.2b,c**). We found no major Watson/Crick strand bias in ribonucleotide distribution throughout the genome (**Figure 2.2a**), and the number of rNMPs identified per nuclear chromosome was proportional to chromosome size (**Figure 2.2d**).

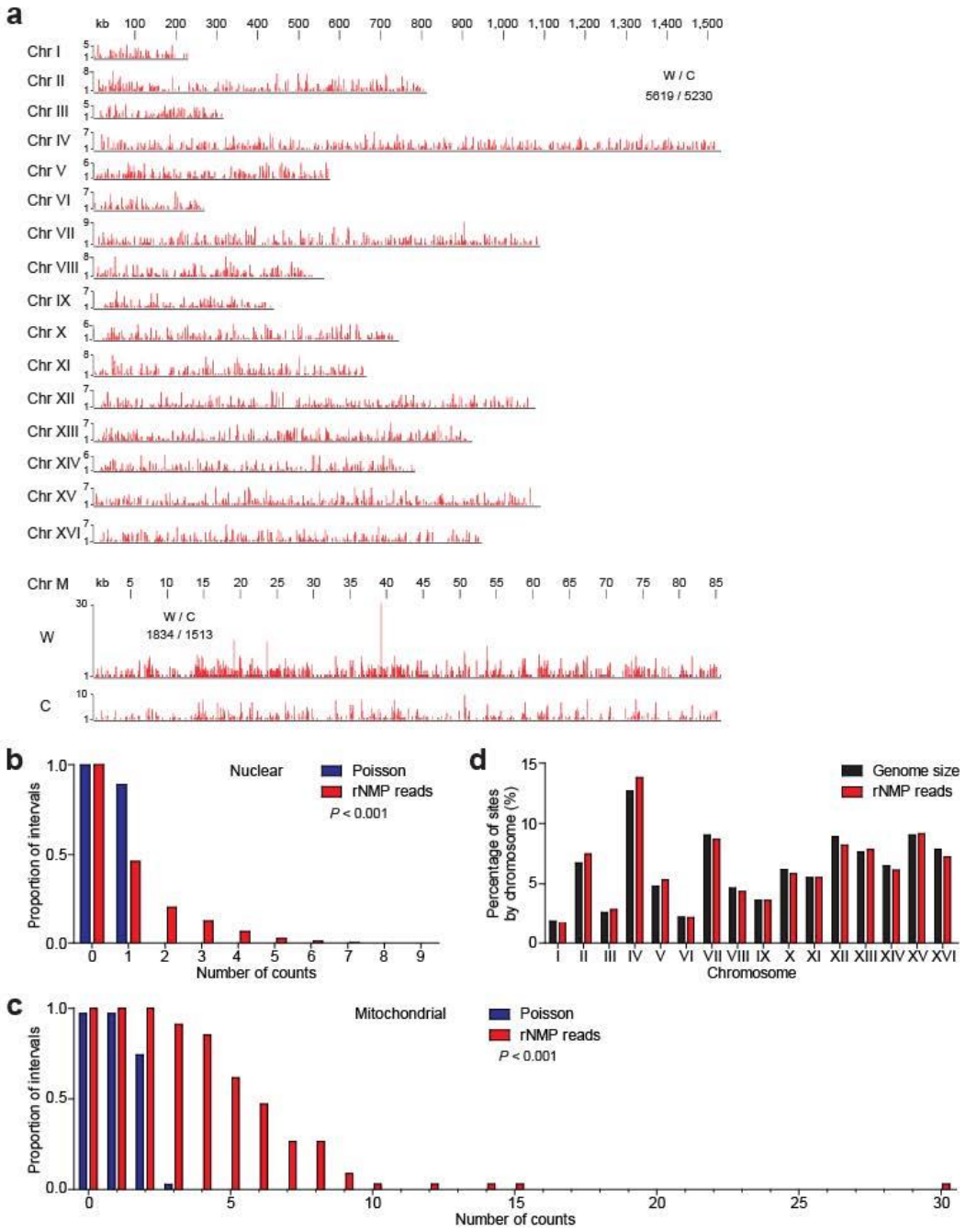


Figure 2.2 Distribution of rNMP incorporation in *S. cerevisiae* genome. (**a**) Ribose-seq map of rNMPs in genomic DNA from *rnh201*Δ (KK-100) cells. The data, as peaks of rNMP reads in red, are shown for the individual nuclear chromosomes (Chr I–XVI) and the two strands of mitochondrial DNA (Chr M). The height of each peak corresponds to the number of reads. A comparison of nuclear and mitochondrial rNMP reads for Watson (W) and Crick (C) strands is also displayed. Raw sequencing reads are available at NCBI GEO [90] under accession GSE61464. The proportion of 2.5 kb windows containing an observed number of rNMPs (red bars) was calculated for nuclear (**b**) and mitochondrial (**c**)

and compared to random expectation based on Poisson frequencies (blue bars). The *P* values (sample sizes of 10,847 and 3347 for nuclear and mitochondrial, respectively) calculated from chi-square goodness of fit test are shown. (**d**) Chromosomal distribution of rNMPs (red bars) is compared to the size of each nuclear chromosome (black bars).

We determined the identity and relative frequencies of incorporated rNMPs, the reverse complement of the 5′ base of each read, as well as flanking bases for the nuclear and mitochondrial genomes in *rnh201*Δ cells. At the site of rNMP incorporation, we found that rCMP and rGMP were incorporated more frequently than expected from the G+C content, while rAMP and in particular rUMP were incorporated less frequently than expected from the A/T content in both nuclear and mitochondrial genomes, indicating a strong bias in the rNMP spectrum considering the A+T rich content of these genomes in yeast (62% and 83%, respectively) (**Figure 2.3a,b**). Looking at the absolute composition of the genomic rNMPs, we found 44% rC, 28.1% rG, 15.4% rA and 12.5% rU in the nuclear genome, and 36.8% rC, 25.6% rA, 19% rG, and 18.7% rU in the mitochondrial genome (**Table A.5**). The difference in the base composition between nuclear and mitochondrial rNMPs is likely due to the higher A+T content of the mitochondrial genome.

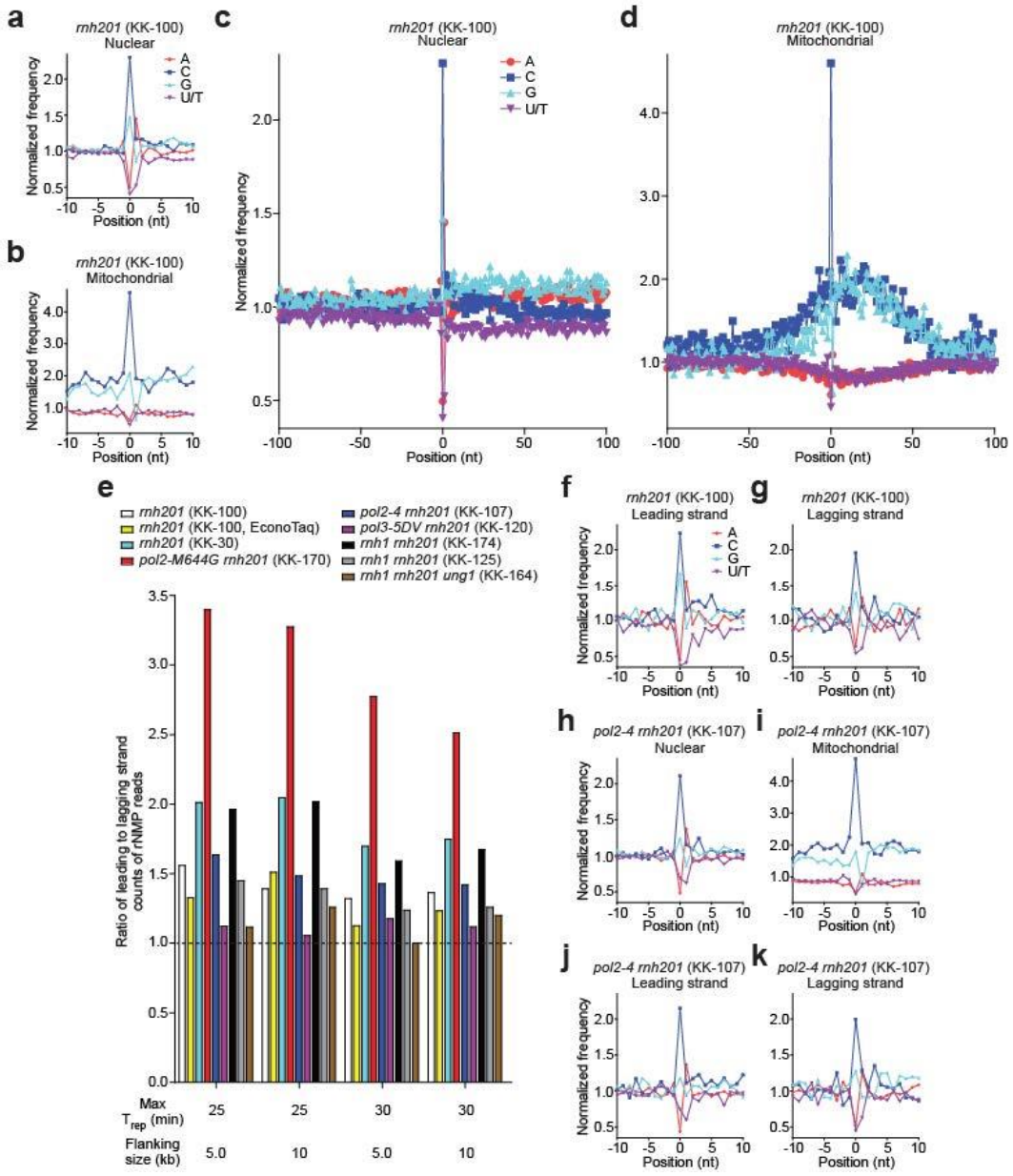


Figure 2.3 Identify and sequence contexts of rNMP incorporation in *S. cerevisiae* genome. Normalized frequency of nucleotides relative to (**a**) nuclear and (**b**) mitochondrial mapped positions of sequences from ribose-seq library of *rnh201*Δ (KK-100) cells. Position 0 is the rNMP. Negative and positive numbers correspond to upstream and downstream positions from the rNMP, respectively. Frequencies were normalized to either nuclear or mitochondrial genomic mononucleotide frequencies. Zoom-out (from position -100 to 100) of normalized frequency of nucleotides for (**c**) nuclear and (**d**) mitochondrial genomes. (**e**) Ratios of rNMPs on leading to lagging strand are shown for all ribose-seq libraries in this study. Early firing ARSs were selected by their replication timing (T_{rep}) while two different flanking sizes were investigated. Horizontal dotted line marks the ratio of 1, when the number of rNMPs on leading strand equals the number of rNMPs on lagging

strand. Normalized frequency of nucleotides relative to mapped positions of sequences in (**f**) leading and (**g**) lagging strands from ribose-seq library of *rnh201*Δ (KK-100) cells. ARSs with T_{rep} of no longer than 25 min were selected with flanking size of 10 kb. Frequencies were normalized to genomic mononucleotide frequencies of either leading or lagging strand of the selected ARSs and flanking size. Normalized frequency of nucleotides relative to (**h**) nuclear and (**i**) mitochondrial mapped positions of sequences from ribose-seq library of *pol2-4 rnh201*Δ (KK-107) cells. Normalized frequency of nucleotides relative to mapped positions of sequences in (**j**) leading and (**k**) lagging strands from ribose-seq library of*pol2-4 rnh201*Δ (KK-107) cells.

The high level of rCMPs and low level of rUMPs observed both for nuclear and mitochondrial DNA in the *rnh201*Δ library are not attributable to differential bypass by the Pfu-based DNA polymerase used for PCR (**Figure A.4** and **Table A.6**), as we also observed similar rNMP patterns using a Taq-based DNA polymerase (**Figures A.5a,b** and **A.6a,b**). Similarly, the nucleotide frequency derived from a ribose-seq library constructed from another *rnh201*Δ strain (KK-30, **Table A.2**) was comparable to that obtained from strain KK-100 both for nuclear and mitochondrial sites (**Figures A.5c,d** and **A.6c,d**). Additional deletion of the gene coding for RNase H1 (*rnh1*Δ), generating *rnh1*Δ *rnh201*Δ strains (KK-174 and KK-125), did not affect the nucleotide frequency of rNMP incorporation (**Figures A.5e-h** and **A.6e-h**). While some variation in the absolute rNMP counts were found among these different libraries in the mitochondrial DNA (**Table A.5**), the high level of rCMP and low level of rUMP remained constant, as well as a preferred rNMP incorporation in G+C rich regions of the mitochondrial DNA. These data support a model in which rNMPs in yeast genomic DNA are present as single, di-, or tri-nucleotides, which are not substrates of RNase H1 [53], and indicate that RNase H1 has minor impact on the distribution of genomic rNMPs.

2.4.3 UNG1 targets uracil from a dUMP but not rUMP embedded in DNA

In order to test whether the low frequency of rUMP incorporation was a consequence of removal by the uracil N-glycosylase, Ung1, we deleted the *UNG1* gene in the RNases H-defective background (*rnh201*Δ *rnh1*Δ *ung1*Δ, strain KK-164) and mapped rNMP sites in these cells. Ung1 repairs dUMP from nuclear and mitochondrial DNA [99]. Whereas Ung1 does not act on uracil in RNA (*e.g.* rRNA) [100], it is not known whether Ung1 can act on rUMP embedded in a DNA duplex. Results presented in **Figures A.5i,j**, **A.6i,j**, and **Table A.5** show that the low level of rUMP incorporation in the chromosomal and mitochondrial genome of a *rnh1*Δ *rnh201*Δ *ung1*Δ strain was similar to rUMP incorporation in a *rnh1*Δ *rnh201*Δ strain, demonstrating that Ung1 does not target uracil in rUMP embedded in DNA.

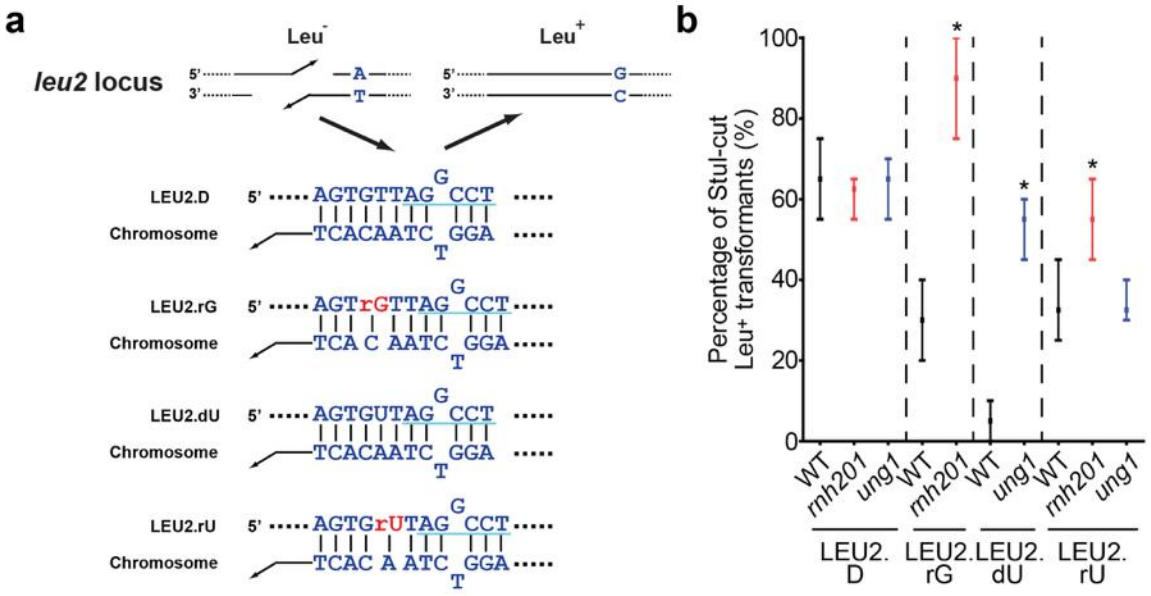


Figure 2.4 Targeting of rGMP and rUMP by RNase H2 and uracil DNA N-glycosylase during DSB repair in *S. cerevisiae* cells. (**a**) Diagram and sequence of the chromosomal *leu2* region targeted by DNA-control LEU2.D, rGMP-containing LEU2.rG, dUMP-containing LEU2.dU, and rUMP-containing LEU2.rU oligos (**Table A.2**). StuI recognition

sequence is underlined in turquoise. Position of either rGMP, dUMP, or rUMP was selected so that it is about 4-5 nt upstream of the G-T mispair. Both RNase H2-initiated excision repair (RER) and base excision repair (BER) remove a short ss DNA region downstream of the damage during the repair [1, 99]. (**b**) The oligos were transformed to either RNase H2- and uracil DNA N-glycosylase-proficient wild-type (WT; FRO-767,768), RNase H2-deficient (*rnh201*; FRO-984,985), or DNA N-glycosylase-deficient (*ung1*; KK-158,159) *S. cerevisiae* cells (see **Table A.1**). Median percentages of StuI-cut Leu^+ transformants from four independent transformations are shown with ranges as bars. For each transformation, 20 Leu^+ transformants were selected for analysis. Mann-Whitney *U*-test was implemented for statistical analysis against the WT. *P* values of less than 0.05 are marked as asterisk. See **Table A.7** for more statistics.

Using a yeast assay of chromosomal double-stranded break (DSB) repair by DNA oligos carrying embedded rGMP, rUMP, or deoxyribose nucleotides only (**Figure 2.4a**), we demonstrate that Ung1 targets uracil from a dUMP but not an rUMP embedded in DNA, while RNase H2 targets only rNMPs (rGMP and rUMP in this experiment), but not dNMPs (**Figure 2.4b** and **Table A.7**). We attribute rNMP incorporation frequencies to the levels of corresponding dNTPs. dCTP and dGTP are typically the least abundant dNTPs [12, 101], and therefore might be depleted faster than dTTP and dATP, increasing the probability of rCMP and rGMP incorporation over rUMP and rAMP. These results are also consistent with the finding that rCMP and rGMP are the most frequently incorporated rNMPs by DNA polymerases *in vitro* under physiological dNTP/rNTP concentrations [18, 63]. This ability of DNA polymerases to incorporate rNMPs into genomic DNA could potentially serve as a mechanism to continue replication under conditions in which one or more dNTP pools are depleted. In the presence of hydroxyurea, a known ribonucleotide reductase inhibitor, higher level of rNMPs are found incorporated in genomic DNA [53]. However, extensive rNMP incorporation would also result in increased breaks and genomic instability.

2.4.4 Pattern of sequences flanking rNMPs in *S. cerevisiae* DNA

Downstream of incorporated rNMPs, we found that the +1 position was most frequently dA and less frequently dG both in the nuclear and in the mitochondrial genomes, with 42 to 52% dA and 6 to 16% dG among all four deoxyribonucleotides (dA, dC, dG, and dT) (**Table A.5**). At the +1 position, dT was also frequent (31 – 40%) in the mitochondrial genome. In mitochondrial DNA, high level of dA or dT at the +1 position 3′ from the rNMP could reflect the high A+T content in the mitochondrial genome. It is also possible that the dA in +1 position influences rNMP incorporation by DNA polymerases. Alternatively, we speculate that dA in the +1 position might stabilize incorporated rNMPs, possibly by affecting base stacking and preventing its repair by non-RER mechanisms. It will be interesting to determine the nearest-neighbor thermodynamic parameters for single rNMPs in DNA duplex and, in particular, the stability trend for the base pair 3′ of the rNMP sites. We recently showed that single rGMPs embedded in a short DNA duplex have a marked effect on the elastic properties of DNA by altering the DNA structure at rNMP sites and the nucleotide 3′ to the rNMPs [37]. Thus, it is reasonable to think that the +1 position 3′ from the rNMP is prone to have altered structure, and it is the most critical site for signaling the presence of an rNMP in DNA because it is the closest nucleotide to the 2′-OH group of the rNMP.

Sites of rNMP incorporation were flanked by sequence contexts that differ between the nuclear and mitochondrial DNA genomes. While nucleotide frequencies up- and downstream of rNMP sites in the nuclear genome were largely similar to background frequencies (**Figure 2.3c**), rNMP sites in mitochondrial DNA were primarily upstream of G+C-rich regions, concentrated in areas in which G+C content was 1.7 to 1.8-fold above

the background (**Figure 2.3d**). Interestingly, mitochondrial G+C tracts have been shown to have recombinogenic properties [102], and mitochondrial DNA recombination has been suggested to initiate mitochondrial DNA replication in yeast [103]. Thus, it is possible that the presence of rNMP sites in yeast mitochondrial G+C clusters influences these recombination events in mitochondrial DNA.

2.4.5 rNMP incorporation by replication DNA polymerases

We next analyzed rNMP incorporation in the leading and lagging strands of yeast nuclear DNA. We selected 154 to 271 early-firing yeast autonomously replicating sequences (ARSs) (activated in the first 25 or 30 minutes, respectively) based on replication timing (T_{rep}) [94]. We examined the type and abundance of rNMPs incorporated in regions 5 or 10 kb upstream and downstream from selected ARSs. This analysis was conducted using all our ribose-seq libraries, including a library derived from yeast RNase H2-deficient cells containing the low fidelity Pol ε mutant (*rnh201Δ pol2-M644G*, **Table A.1**). Because yeast Pol ε is mainly responsible for leading strand synthesis during DNA replication, yeast cells containing the *pol2-M644G* mutation, which leads to increased rNMP incorporation, would be predicted to contain more rNMPs on the leading than on the lagging strand [41]. We found higher rNMP incorporation on the leading strand of DNA replication (**Figure 2.3e**), consistent with previous observations that the leading strand DNA Pol ε incorporates more rNMPs than the lagging strand Pol δ [12]. As expected, analysis of rNMPs from *rnh201Δ pol2-M644G* cells revealed a stronger bias of rNMP incorporation on the leading strand compared to all other libraries (**Figure 2.3e**). However, the increase in rNMP incorporation by the low fidelity Pol ε mutant did not change the overall rNMP spectrum, having similar

patterns of rNMP incorporation to libraries derived from wild-type Pol ε strains **(Figures A.5k,l** and **A.6k,l**). Furthermore, we examined whether the spectrum of rNMP incorporation was different between the leading and the lagging strand of DNA replication. Cells containing either *rnh201*Δ or *rnh1*Δ *rnh201*Δ mutations had a similar spectrum of rNMP incorporation on the leading and the lagging strand (**Figure 2.3f,g** and **Figure A.7a,b**). In contrast, mapping of rNMPs in yeast cells carrying a mutant allele of DNA Pol ε that is defective in proofreading activity (*pol2-4*, **Table A.1**), showed a lower frequency of rA versus rU in the nuclear but not in the mitochondrial genome (**Figure 2.3h,i**; **Figure A.6m,n**; and **Table A.5**), and showed a bias for lower rA than rU only on the leading strand (**Figure 2.3j,k**), which was not observed in libraries derived from wild-type nor low fidelity mutant Pol ε strains. The rNMP spectra for a strain with proofreading-defective DNA Pol δ, (*pol3-5DV*, **Table A.1**) were not different from those containing the wild-type Pol δ (**Figures A.5m,n**; **A.6o,p**; **A.7c,d**; and **Table A.5**). These results suggest that DNA Pol ε can proofread rNMPs in DNA, in particular rUMPs, and that this activity is superior to DNA Pol δ, consistent with previous biochemical studies [18, 63].

2.4.6 Hotspots of rNMP incorporation in *S. cerevisiae* genome

We performed two types of analyses to determine potential hotspots of rNMP incorporation in *S. cerevisiae* genome. We identify enriched regions of rNMP incorporation in genomic DNA from ribose-seq data (see **Materials and Methods**). We found several regions of notable rNMP incorporation in mitochondrial DNA for each ribose-seq library in this study, and a few of these regions are displayed in **Figure 2.5a**. Because this analysis excludes all reads aligning to more than one position in the genome, a second analysis was

performed with respect to specific loci in order to identify single-nucleotide hotspots that were reproducibly present in multiple ribose-seq libraries (see **Materials and Methods**).

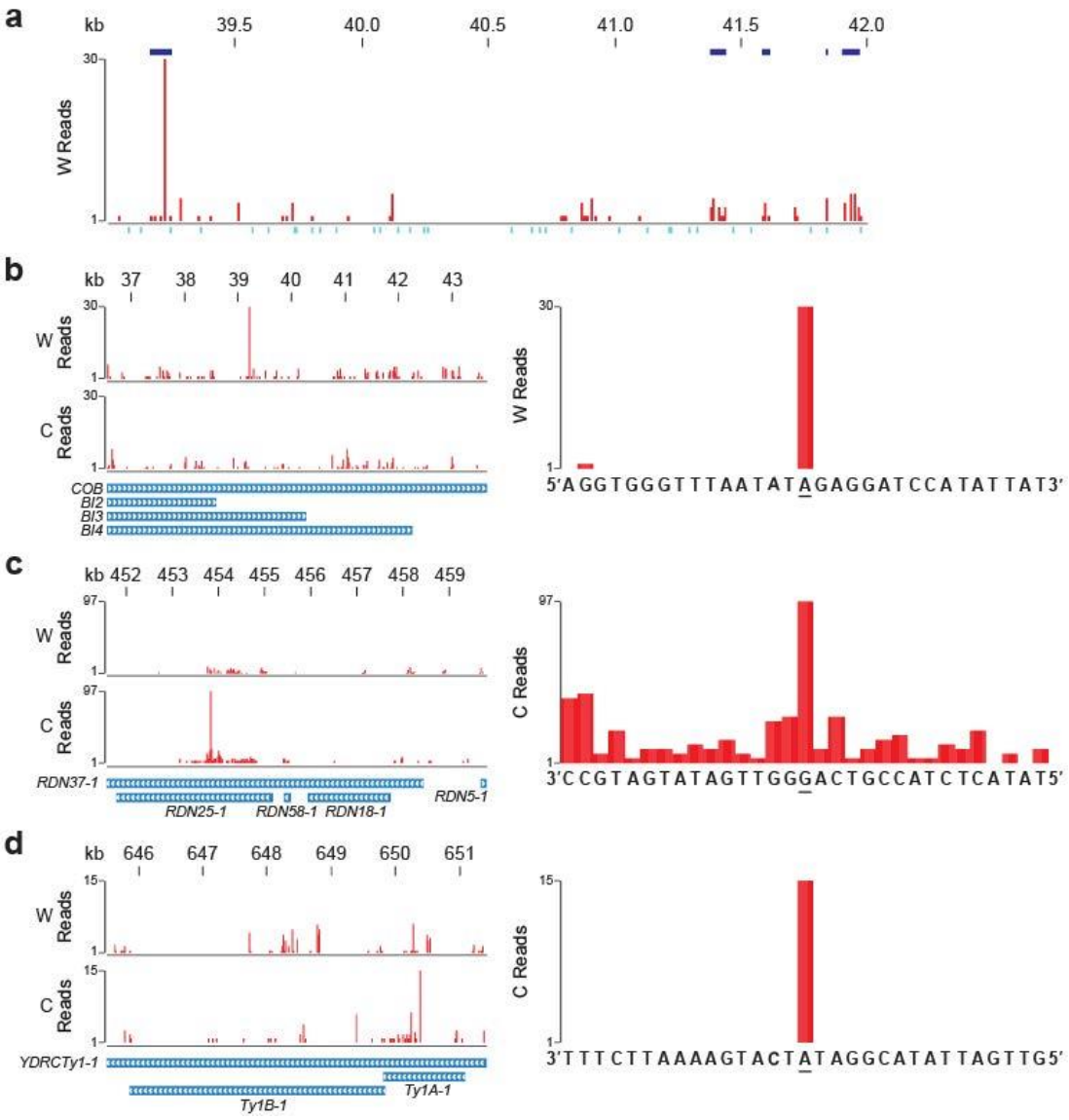


Figure 2.5 Hotspots of rNMP incorporation within *S. cerevisiae* mitochondrial DNA, rDNA repeat, and *Ty1*. (**a**) Ribose-seq map of rNMPs in a 3 kb region (39,001–42,000) in mitochondrial DNA from *rnh201*Δ (KK-100) cells with enriched regions of rNMP incorporation. Peaks of rNMP incorporation are shown in red for the Watson (W) strand. Enriched regions of rNMPs with *q*-value of less than 0.001 are shown in blue. Position of restriction sites of enzymes used for genomic fragmentation is displayed in turquoise. (**b**) Ribose-seq map of rNMPs in *COB* locus of mitochondrial DNA from *rnh201*Δ (KK-100) cells (left). Zoom-in map (right) with sequence at the hotspot site (underlined). (**c**) Ribose-seq map of rNMPs in rDNA repeat locus on Chr XII of genomic DNA from *rnh201*Δ (KK-

100) cells (left). Because of the presence of two rDNA repeats on Chr XII in the reference genome (sacCer2), only the first repeat unit is shown as an example. Multiple-alignment data are shown. Zoom-in map (right) of the rDNA hotspot. (**d**) Ribose-seq map of rNMPs in *Ty1* locus of genomic DNA from *rnh201*Δ (KK-100) cells (left). Because of the presence of multiple copies of *Ty1* in the genome, *YDRCTy1-1* on Chr IV is shown as an example. Multiple-alignment data are shown. Zoom-in map (right) of the *Ty1* hotspot. The data, as peaks of rNMP reads in red, are shown for the Watson (W) and Crick (C) strands.

We identified hotspots of rNMP incorporation within sequences present in multiple copies per yeast cell, *i.e*. the mitochondrial genome (~80 copies) [104], the ribosomal DNA (rDNA) repeats (~140) clustered on chromosome XII [105], and the yeast retrotransposon (Ty), of which there are ~30 copies encoded on multiple chromosomes [106]. In mitochondrial DNA, we found a marked hotspot at an rAMP on the Watson strand in the cytochrome oxidase B gene (*COB*) and the overlapping maturase *BI3* and *BI4* genes (**Figure 2.5b** and **Table A.8**), in addition to several other hotspots (**Table A.8**). In the rDNA locus, the strongest hotspot was found in gene *RDN37-1* and the overlapping *RDN25-1* at an rGMP (**Figure 2.5c** and **Table A.8**). In the yeast *Ty1* sequence, we found a hotspot at an rAMP in the coding sequence of *TY1A-1* (**Figure 2.5d** and **Table A.8**).

The occurrence of such hotspots indicates that there are preferred sites for rNMP incorporation in the mitochondrial genome, rDNA, and *Ty1* sequences. In addition to the recombinogenic properties of mitochondrial G+C clusters discussed above, yeast rDNA and Ty are also active in recombination [105, 107]. Frequent rNMP incorporation could trigger recombination, similar to rNMPs embedded in the mating type locus of *Schizosaccharomyces pombe* [73]. The rNMPs detected in Ty DNA could originate from cDNA rather than genomic DNA. Because *rnh201 rnh1* cells have abundant Ty cDNA

[108], if rNMPs are incorporated in *Ty1* during the process of reverse transcription, which forms the cDNA, we would expect a different rNMP pattern in *Ty1* DNA in *rnh201*Δ *rnh1*Δ vs. *rnh201*Δ cells. Although we did not observe major differences in the rNMP spectra derived from *rnh201*Δ single mutant vs. *rnh201*Δ *rnh1*Δ cells at the *Ty1* locus or in general (**Table A.8**; **Figure 2.3e**; **Figures A.5**; and **A.6**), it would be of interest to conduct *in vitro* tests to determine whether Ty reverse transcriptase incorporates rNMPs frequently opposite to RNA and/or DNA, and whether it has a particular bias for rNMP incorporation.

2.5 Discussion

rNMP incorporation has been extensively studied in recent years; however, determining the identity and locating sites of rNMP incorporation in genomic DNA has not yet been possible. We describe the development of ribose-seq to map sites of rNMPs in genomic DNA. Our ribose-seq approach specifically targets and captures rNMPs embedded in DNA via alkaline hydrolysis, resulting in DNA fragments with 2′,3′-cyclic phosphates or 2′-phosphates which are subsequently ligated to adaptor sequences by AtRNL. Properties of alkaline cleavage of rNMPs and AtRNL ligation exclude Okazaki fragments and DNA abasic sites from our ribose-seq library, allowing construction of libraries containing stably incorporated rNMP sites and analysis by high-throughput DNA sequencing. Our ribose-seq enabled the determination of the widespread but nonrandom distribution of rNMPs in budding yeast genomic nuclear and mitochondrial DNA with several hotspots. Base preference and strand bias relevant to replication of rNMP incorporation in yeast genomic DNA were found while flanking DNA sequences were also found to be a substantial factor for incorporation of rNMPs. Our findings with ribose-seq show both validation of the approach and determination of new aspects of rNMP incorporation in the yeast genome. It

would be interesting to determine the rNMP spectrum in cells with defects in rNMP removal pathways other than RER, either in RNase H2 wild-type or null cells growing under normal and/or stress conditions. Ribose-seq serves as foundation for such studies and as novel technology to better understand the genetic and epigenetic consequences of rNMPs in DNA, and their impact on the structure and function of DNA and chromatin. Moreover, specific signatures of rNMP incorporation may represent novel biomarkers for human diseases, such as AGS, cancer, and other degenerative disorders.

2.6 Acknowledgements

We thank N.V. Hud and L.D. Williams for support with urea-PAGE gels and for advice on this study and the manuscript; M. Goodman, B. Weiss and I.K. Jordan for suggestions on this study and the manuscript and assistance with data analysis; S. Garrey for AtRNL and Tpt1 protein purification; C. Cox for sequencing; Y. Shen for assistance on statistical analysis; A. Gombolay and L. Shetty for technical help; and all members of the Storici laboratory for advice in the course of the study. This research was supported by the National Science Foundation award number MCB-1021763 (to F.S.), the Georgia Research Alliance award number R9028 (to F.S.), an American Cancer Society Research Scholar Grant (to J.R.H.), a Damon Runyon-Rachleff Innovation Award from the Damon Runyon Cancer Research Foundation (to J.R.H.), and the University of Colorado Golfers Against Cancer (to J.R.H.).

CHAPTER 3. ABASIC AND OXIDIZED RIBONUCLEOTIDES EMBEDDED IN DNA ARE PROCESSED BY HUMAN APE1 AND NOT RNASE H2

The study in Chapter 3 consists of the work published in *Nucleic Acid Research* Vol. 45, No. 19 (2017) 11193-11212. All the RNase H2 assays in Figures 3.3 and B.3, and APE1 experiments in Figures 3.4 and 3.6, along with initial APE1 enzymatic experiments on rNMPs and abasic rNMPs were my contributions to this study.

Malfatti, M.C.[1†], Balachander, S.[2†], Antoniali, G.[1], Koh, K.D.[2,3], Saint-Pierre, C.[4], Gasparutto, D.[4], Chon, H.[5], Crouch, R.[5], Storici, F.[2], & Tell G.[1]

[1]Department of Medicine, University of Udine, Udine, Italy.

[2]School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA.

[3]Currently at the School of Medicine, University of California, San Francisco, USA.

[4]Chimie Reconnaissance & Etude Assemblages Biologiques, Université Grenoble Alpes, Grenoble, France.

[5]Developmental Biology Division, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institute of Health, Bethesda, Maryland, USA.

[†]Equal contribution

3.1 Abstract

Ribonucleotide 5'- monophosphates (rNMPs) are the most common non-standard nucleotides found in DNA of eukaryotic cells, with over 100 million rNMPs transiently incorporated in the mammalian genome per cell cycle. Human ribonuclease (RNase) H2 is the principal enzyme able to cleave rNMPs in DNA. However, whether RNase H2 may process abasic or oxidized rNMPs incorporated in DNA is unknown. The base excision repair (BER) pathway is mainly responsible for repairing oxidized and abasic sites into DNA. Here we show that human RNase H2 is unable to process an abasic rNMP (rAP site) or a ribose 8oxoG (r8oxoG) site embedded in DNA. On the contrary, we found that recombinant purified human apurinic/apyrimidinic endonuclease-1 (APE1) and APE1 from human cell extracts efficiently process a rAP site in DNA and have weak endoribonuclease and 3'-exonuclease activities on r8oxoG substrate. Using biochemical assays, our results provide the first evidence of a human enzyme able to recognize and process abasic and oxidized ribonucleotides embedded in DNA.

3.2 Introduction

Incorporation of ribonucleotides monophosphate (rNMPs) in DNA is a frequent phenomenon, which is considered the most common type of 'DNA damage' occurring in normal cells [6, 70]. Ribose-seq and other approaches recently developed for mapping sites of rNMPs in DNA have shown widespread but not random distribution of rNMPs in chromosomal DNA of budding and fission yeast [68-71]. The number of rNMPs identified per nuclear chromosome was found to be proportional to chromosome size, and

quantitation approaches have estimated a 600 thousands rNMPs in budding yeast genome and over 100 millions in mouse genome [27].

The incorporation of rNMPs in genomic DNA may be due to: *i*) the disequilibrium in the cellular pool of deoxyribonucleotides (dNTPs) and ribonucleotides (rNTPs) [70]; *ii*) an incomplete elimination of RNA primers used in the generation of Okazaki fragments [18], *iii*) an oxidation of the deoxyribose sugar into ribose[24], and last but not least, *iv*) an imprecise 3'-exonucleolytic proofreading activity of replicative DNA polymerases, which do not discriminate rNMPs from dNMPs pool [12, 18, 70, 109]. Furthermore, taking into consideration all rNMPs that are synthesized during lagging strand synthesis, more than 100 million rNMPs are introduced into the mammalian genome per cell cycle [27]. In addition, it has been estimated that the amount of rNTPs is generally 40–350 - fold higher than that of dNTPs in cycling cells [12, 13, 109] increasing the probability of incorrect rNMP incorporation during DNA replication and repair [12, 18, 70, 109].

The effects of the 2′-hydroxyl group of the ribose sugar within an rNMP embedded in DNA are numerous: it mainly alters DNA elasticity and structure in a sequence dependent manner [2, 37, 110, 111] and affects the activity and function of several DNA-interacting proteins, in addition to increase the DNA fragility and mutability [2, 6]. Moreover, rNMPs in DNA can be template for DNA synthesis [30, 31], although the DNA polymerases processivity on rNMP tracts is reduced [31].

The main pathway deputed to rNMPs removal is ribonucleotide excision repair (RER), in which the principal enzyme is RNase H2. It was demonstrated how RNase H2 deficiency in mammalian cells is associated to DNA damage repair activation and, in humans, to

pathology. RNase H2-null murine embryonic fibroblasts (MEFs) activate a p53-dependent damage response, whereas null-RNase H2 are embryonic lethal [44]. In humans, mutations in each of the three subunits of RNase H2 are associated with the neurological syndrome of Aicardi-Goutières (AGS), which causes severe brain dysfunction [54, 112, 113]. Altered RNase H2 function in AGS patients may result in increased level of rNMPs in DNA, which could in turn activate DNA damage response signaling and induce innate immune response [53, 114].

Besides RER, other different DNA repair mechanisms are active to remove rNMPs embedded in DNA [27, 35]. In the absence of RNase H2, topoisomerase I cleavage [35], followed by nick processing by Srs2–Exo1, can remove some rNMPs [50, 56]. rNMPs in DNA can be also targeted by the nucleotide excision repair (NER) factors in bacteria [115]. However, *in vitro* experiments showed that human NER proteins are not active to remove rNMPs embedded in DNA [61]. Differently, the mismatch repair (MMR) mechanism targets mismatches with rNMPs both in *E. coli* and *S. cerevisiae* genomic DNA [29]. Until now, there is no proof that the base excision repair (BER) mechanism plays any role in removing rNMPs from DNA.

BER is known to repair a wide spectrum of oxidative lesions in nuclear and mitochondrial DNA [64, 65], and preventing cancer [116, 117]. Abasic sites, which form by spontaneous hydrolysis of the N-glycosidic bond in DNA or following removal of a damaged base by BER glycosylases, are major targets or intermediate substrates in the BER pathway of DNA repair [8]. It has been estimated that up to 10,000 abasic sites are formed per human genome per day [5]. Despite the fact that spontaneous depurination occurs ~1,000 times slower in RNA than DNA [25], due to the high abundance of rNMPs in genomes, with

more than 100 million rNMPs transiently present in mammalian DNA during one replication cycle, considering rNMP incorporation by DNA polymerases during DNA replication and repair, and RNA primers of Okazaki fragments [27], the possibility that abasic and oxidized rNMPs (such as 7,8-Dihydro-8-oxo-riboguanosine) are present in DNA and are targets of BER is quite real and worth careful study. These data, together with other recent findings about the ability of *S. pombe* Pol 4, *M. smegmatis* DinB2 and human Pol β to insert and elongate oxidized rGMP when paired with dA during DNA replication [118-120], underscore the necessity to determine how cells can target and remove oxidized rNMPs or rAP sites from DNA.

Because recent studies point towards a new function of BER in RNA surveillance [66, 67], there is high likelihood that BER could be involved in the processing of rNMPs in DNA, particularly in the case of chemically modified rNMPs, such as abasic and oxidized rNMPs. Identifying whether BER may target normal and modified rNMPs in DNA is important to better understanding the mechanism of genotoxicity of reactive oxygen species, the function and the impact of BER defects in human disease and cancer mechanisms. In the absence of proper repair mechanisms to cope with these kind of lesions, even a single or a few modified rNMPs present in a genome per cell cycle could lead to mutations and/or genomic rearrangements.

Findings from our and other laboratories have revealed an important involvement of the apurinic/apyrimidinic endonuclease 1 (APE1) in RNA metabolism and RNA-decay [66, 67, 121, 122]. APE1 is by far one of the most studied enzymes in the BER pathway for its altered expression in different human pathologies ranging from neurodegenerative to cancer disorders [123]. Its role in DNA repair is primarily due to its ability to act as an

endonuclease, specifically being able to cleave 5′ to deoxy- abasic sites, which results in a strand break with 3′-hydroxyl and 5′-phosphodeoxyribose termini. APE1 also has redox activity needed to modulate the DNA binding ability of several transcription factors [123]. As recently demonstrated, APE1 can endonucleolytically cleave abasic single-stranded RNA [66, 67, 122], has a 3′-RNA phosphatase activity, and a weak 3′-5′ exoribonuclease activity [124]. Moreover, it has been demonstrated that APE1 has nucleotide incision repair (NIR) activity on modified bases, such as 5,6-dihydro-2'-deoxyuridine, 5,6-dihydrothymidine, 5-hydroxy-2'-deoxyuridine, 5-hydroxycytosine, which are directly repaired by APE1 bypassing the action of specific glycosylases [125-127]. Therefore, we hypothesized that APE1 can be involved in processing rNMPs in DNA, particularly in the case of abasic and oxidized rNMPs.

Here, we found that eukaryotic RNase H2 from yeast, mouse and human is inactive on a rAP site in DNA in different assays. We discovered and characterized an unknown APE1 activity on abasic ribonucleotide embedded in DNA. We then compared the ability of human RNase H2 to cleave at an oxidized ribonucleotide (r8oxoG) incorporated in a DNA substrate and analyzed the activities of 8-oxoguanine DNA glycosylase 1 (OGG1) and APE1 to recognize and cleave this particular type of damage. Our data demonstrate that APE1, but not human RNase H2 and OGG1, has a weak endoribonuclease activity on the oxidized substrate.

3.3 Materials and Methods

3.3.1 Double strand oligonucleotide description and annealing sequence

All oligonucleotides and their complementary sequences used in this study are listed in **Table B.1** (see **Figure B.1**). ss_dG_40 oligonucleotide and its reverse complementary sequence, ss_dC_40, were purchased from Invitrogen (Grand Island, NY, USA). ss_rG_40 oligonucleotide and the DNA oligo containing a tetrahydrofuran abasic deoxyribonucleotide, ss_dF_40, or tetrahydrofuran abasic ribonucleotide, ss_rF_40, as well as ribo- 1'OH abasic containing oligonucleotide, ss_rOH, were purchased from Dharmacon (GE Healthcare, Lafayette, CO, USA). The 26-mer oligonucleotide containing a tetrahydrofuran, ss_dF, and its reverse complementary sequence ss_dC, were synthesized from Metabion International AG (Steinkirchen, Germany). The 25-mer dG-, rG-, d8oxoG-containing oligonucleotides and complementary oligonucleotides were synthesized from Metabion International AG (Steinkirchen, Germany).

ss_dg_40, ss_rG_40, ss_dF_40 and ss_rF_40 oligonucleotides were 5′ end-labeled with [γ-^{32}P] ATP (PerkinElmer, Boston, MA, USA) by T4 polynucleotide kinase (PNK) (New England BioLabs, Ipswich, MA, USA) in a reaction mixture containing 10 μM ATP using 10X PNK buffer (New England BioLabs, Ipswich, MA, USA). This labeling reaction was incubated at 37 ℃ for 1 hour, followed by inactivation at 65 ℃ for 20 minutes. The reactions were purified by using Illustra MicroSpin G-25 column (GE Healthcare, Buckinghamshire, UK).

The remaining remaining oligonucleotides were labelled with either IRDye700, IRDye800 fluorophores or Cyanine5 at 5' end, as specified in **Table A.1**, purified through RP-HPLC, checked in Mass Check and re-suspended in RNase- and DNase- free water.

Synthesis of oligonucleotide containing an internal ribose 8-oxo-guanosine (r8oxoG) and an IRDye700 fluorophore at 5' end was in-house carried out on an Applied Biosystems 392 DNA/RNA synthesizer using the phosphoramidite chemistry, associated with the phenoxyacetyl protecting group for the nucleobases and the *tertio*-butyldimethylsilyle protecting group at the 2'-OH position of the ribonucleoside residue [128]. Upon completion, the oligonucleotide was de-protected in concentrated aqueous ammonia for 6 hours at 55°C, followed by a desilylation step with triethylamine trihydrofluoride (8 hours at room temperature) [128] and was finally purified by preparative 20% denaturing PAGE using UV-shadowing detection. After desalting by size exclusion, the r8oxoG oligonucleotide was quantified by UV measurements at 260 nm and its purity was checked by RP-HPLC analysis together with MALDI-TOF mass measurements (**Figure B.5a,b**). Sample was then lyophilized and frozen at -20˚C until use.

All oligonucleotides oligonucleotides used in the present study were re-suspended in RNase- and DNase- free water at 100 µM. 100 pmol of each oligonucleotide was annealed with an excess of 150 pmol of its complementary DNA oligonucleotide (as indicated in **Table B.1**) in 10 mM TrisHCl pH 7.4 and 10 mM $MgCl_2$, heated at 95°C and cooling down over night in the dark.

3.3.2 Plasmid and expression of recombinant proteins

Plasmid and expression of human recombinant OGG1 enzyme was purified as described in Audebert *et al*., [129]. Plasmids and expression of human recombinant APE1 wild type (WT) and respective mutants (APE1 NΔ33 and APE1 E96A) were produced as explained in Fantini *et al.*, [130] and in Erzberger and Wilson [131]. Plasmid and expression of yeast, mouse and human recombinant RNase H2 were produced as explained in Chon *et al.*, [47, 132].

3.3.3 Cell lines and silencing experiments

HeLa cells (human cervical carcinoma) (ATCC®, Milan, Italy) were grown in DMEM (EuroClone, Milan, Italy) supplemented with 10% fetal bovin serum (FBS-EuroClone, Milan, Italy), penicillin (100 U/ml), streptomycin (100 mg/ml) and l-glutamine (2 mM) (EuroClone, Milan, Italy) and cultured in a humidified incubator at 5% CO_2 at 37°C. For silencing experiments, 15×10^4 cells were seeded and transfected with 5' UACUCCAGUCGUACCAGACCU 3' siAPE1 (100 pmol) or siGENOME SMART pool siRNase H2A (50 pmol) or 5’ CCA UGA GGU CAG CAU GGU CUG UU 3’ scramble control siRNA (100 pmol) (GE Dharmacon, Milan, Italy) by using Oligofectamine™ Reagent (GE Dharmacon, Milan, Italy) as per manufacturer’s indications. After 72 hours upon transfection, cells were harvested by trypsinization and centrifuged at 250 × g for 5 minutes at 4°C. Supernatant was removed, and pellet was washed once with ice-cold phosphate-buffered saline without Calcium and Magnesium (PBS-Euroclone, Milan, Italy) and then centrifuged again (250 × g for 5 minutes at 4°C).

3.3.4 Preparation of nuclear cell extracts (NCE)

After washing with PBS, cells were collected in cold PBS added with 0.1 M DTT and 0.5 mM phenylmethylsulfonyl fluoride (PMSF). Cells were centrifuged at 800 × g for 10 minutes at 4°C and the supernatant was removed. Pellet was re-suspended in a cold hypotonic solution containing 10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM $MgCl_2$, 0.1 mM EDTA pH 8.0 complemented with 0.1 mM DTT, 0.5 mM PMSF, 1 mM protease inhibitor (PI), 1 mM NaF, 1 mM Na_3VO_4. After centrifugation at 800 × g for 10 minutes at 4°C, cytosolic proteins (CCE) were collected whereas intact nuclei were pelleted. Pellet was washed to discard any contamination from cytosol and it was subsequently re-suspended with a cold hypertonic solution 20 mM HEPES pH 7.9, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.1 mM EDTA pH 8.0, 5% glycerol complemented with 0.1 mM DTT, 0.5 mM PMSF, 1 mM PI, 1 mM NaF, 1 mM Na_3VO_4 and incubated on ice for 30 minutes. At the end, the sample was centrifuged at 15,000 × g for 20 minutes at 4°C and collected the supernatant containing nuclear proteins (NCE). Quantification of each sample was performed by colorimetric Bradford assays (Bio-Rad, Milan, Italy).

3.3.5 Preparation of whole cell extracts (WCE)

After washing with PBS, cells were harvested by trypsinization and centrifuged at 250 × g for 5 minutes at 4°C. The supernatant was removed, washed once with PBS and centrifuged again. Pellet was re-suspended in a lysis solution containing 50 mM Tris HCl (pH 7.4), 150 mM NaCl, 1 mM EDTA, 1% wt/vol Triton X-100 supplemented with 1 mM PI, 1 mM DTT, 0.5 mM PMSF, 1 mM NaF and 1 mM Na_3VO_4. After centrifugation at 15,000 × g for 20 minutes at 4°C, the supernatant is considered as whole cell extract (WCE). Proteins

of each sample were quantified using a colorimetric Bradford assays (Bio-Rad, Milan, Italy).

3.3.6 Enzymatic activity assays

To measure enzymatic activity of recombinant proteins and NCE on different substrates, each reaction was prepared following doses, time points and buffers specified in detail into the legend of each experiment. Final volume for each reaction was 10 µl. At the end of all reactions, samples were blocked with a stop solution, containing 99.5% v/v Formamide (SIGMA-ALDRICH, Milan, Italy) supplemented with 10X Orange Loading Dye (LI-COR Biosciences, Milan, Italy) and heated at 95°C for 5 minutes. Then, all samples were loaded onto a 7 M denaturing 20% polyacrylamide gel in TBE buffer pH 8.0 and run at 4°C at 300V for 1 hour. Then, the gel was visualized with an Odyssey CLx Infrared Imaging system (LI-COR GmbH, Germany). The signals of the non-incised substrate (S) and the incision product (P) bands were quantified using Image Studio software (LI-COR GmbH, Germany). When using the ds_rOH:dC and ds_r8oxoG:dC oligonucleotides, a very small amount of cleavage product was seen in samples not treated with recombinant proteins and/or cell extracts due to the reactivity of this molecule, which was spontaneously degraded. During the analysis, this band has been always subtracted from bands obtained following treatment with recombinant proteins and/or extracts.

For radioactive experiments, reactions were stopped by adding 2X denaturing PAGE gel buffer (0.01% bromophenol blue, 95% formamide and 20 mM EDTA pH 8.0) and heating to 95 °C for 5 min. After dilutions, the products were analyzed by 15% (wt/vol) polyacrylamide, 8 M urea gel electrophorersis (urea-PAGE). 20-100 Oligonucleotide

Length Standard (Integrated Device Technology, Coralville, IA, USA) was used as a ladder (M). After electrophoresis, gels were exposed to phosphor screen overnight. Images were taken with Typhoon Trio+ (GE Healthcare, Lafayette, CO, USA) and obtained with ImageQuant (GE Healthcare). Band intensities were quantified by Multi Gauge V3.0 (Fujifilm).

3.3.7 Electrophoretic mobility shift assay analysis (EMSA)

Proteins binding to nucleic acids was assessed by EMSA analysis as already described by Fantini *et al*., [130]. Briefly, the indicated amounts of recombinant purified proteins or cell extracts were co-incubated with 250 fmol of the probe (25 nM) at 37°C for 30 minutes. Reactions were prepared in a buffer containing 8 mM HEPES, 10 mM KCl, 400 µM EDTA pH 8.0, 5 mM DTT and 2% glycerol in a 10 µl final volume. Moreover, salmon sperm DNA (SSD) (SIGMA-ALDRICH, Milan, Italy) was added like as DNA competitor. Samples were loaded onto an 8% w/vol native polyacrylamide gel in Tris-Sodium Acetate-EDTA pH 8.0 (TAE) buffer and run at 4°C at 150V for 1 hour followed by 3 hours at 250V.

3.3.8 Statistical analysis

Statistical analyses were performed by using the Student's t test. $P<0.05$ was considered as statistically significant.

3.4 Results

3.4.1 Human, yeast or mouse RNase H2 does not process an rAP site in DNA

RNase H2 is the principal protein able to process paired and mismatched rNMP sites embedded in DNA by generating a nick to their 5′ side [29]. Up to now, whether RNase H2 can cleave a rAP site incorporated in a duplex DNA is unknown. In order to test this hypothesis, we measured RNase H2 ability to cleave a modified 25-mer DNA oligonucleotide, called ds_rOH:dC, in which a 1'-OH abasic rNMP was incorporated into a DNA substrate as shown in **Figure 3.1a** (see also **Table B.1** and **Figure B.1**).

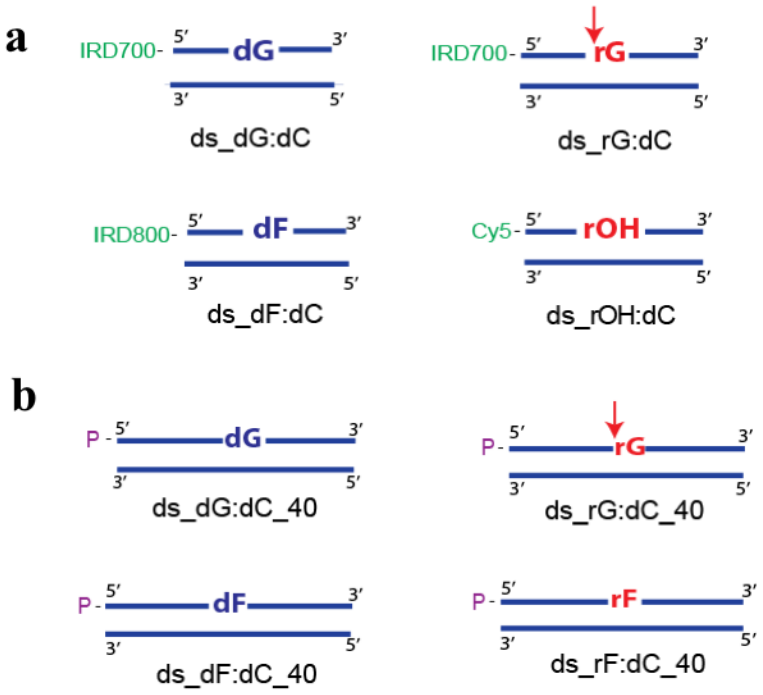


Figure 3.1 Scheme of substrates used to test cleavage of an rAP site in DNA. DNA nucleotides are in blue, RNA in red. The 5′ and 3′ ends of each DNA strand are indicated. (**a**) Scheme of double strand (ds) DNA 25-mer substrates (single strand (ss) oligonucleotide annealed to ss_dC containing-complementary oligonucleotide) containing a dGMP, rGMP,

dF and 1′ OH abasic rNMP (rOH) site in the 13th position. The 'IRD700' in green indicate IRDye 700 phosphoramidite dye tagged at the 5' end. The 'IRD800' in green indicate IRDye 800 phosphoramidite dye tagged at the 5' end. The 'Cy5' in green indicates cyanine dye tagged at the 5′-end of the top strand of the duplex. (**b**) Scheme of the ds_DNA 40-mer substrate (ss oligonucleotide annealed to its ss complementary oligonucleotide) containing a dGMP, rGMP, dF and rF site in the 20th position. The P in purple indicates radiolabelled ^{32}P at the 5′-end of the top strand of the duplex. The red arrow indicates the cleavage position by RNase H2 5' to the rGMP site.

Recombinant human RNase H2, composed of its three subunits, was purified as explained in **Materials and Methods** (**Figure B.2**) and its activity was tested on ds_rOH:dC oligonucleotide in parallel with dG- and rG- containing oligonucleotides as negative and positive controls, respectively. As reported in **Figure 3.2**, the enzyme had no activity on ds_dG:dC and ds_dF:dC (containing a tetrahydrofuran residue mimicking the abasic site) oligonucleotides, whereas it efficiently cleaved the canonical rG substrate as expected. In addition, we tested the mismatched rG- containing oligonucleotide, ds_rG:dA, confirming that RNase H2 protein is also active on this type of substrate (**Figure 3.2a**). On the other hand, no activity was detectable on ds_rOH:dC demonstrating that recombinant human RNase H2 is not able to process an abasic rNMP embedded in DNA. To confirm these data in cells, we tested the activity of RNase H2 protein obtained from nuclear cell extracts on the same substrates, as above (**Figure 3.2b,c** and **Figure B.3b**). To this aim, RNase H2 expression was downregulated in HeLa cells through specific siRNA and the endoribonuclease activities of nuclear extracts from knocked down and control cells were then assayed. Western blotting analyses performed on nuclear extracts from control (Scramble) and knocked down (siRNase H2) cells demonstrated the efficiency of RNase H2 downregulation (about 50%) upon transfection with specific siRNA sequences (**Figure**

B.3a). We incubated Scramble or siRNase H2 cell extracts with different substrates for the indicated time points. Following knock-down of RNase H2, we found a decreased cleavage of ds_rG:dC, as expected (**Figure 3.2c**). Surprisingly, we found that ds_rOH:dC was also cleaved; however, ds_rOH:dC cleavage was completely unaffected by siRNase H2 cell extracts (**Figure 3.2c**). These data suggest that while human RNase H2 is inactive on an abasic rNMP embedded in DNA, there is another enzyme/s capable of cleaving it in human cells.

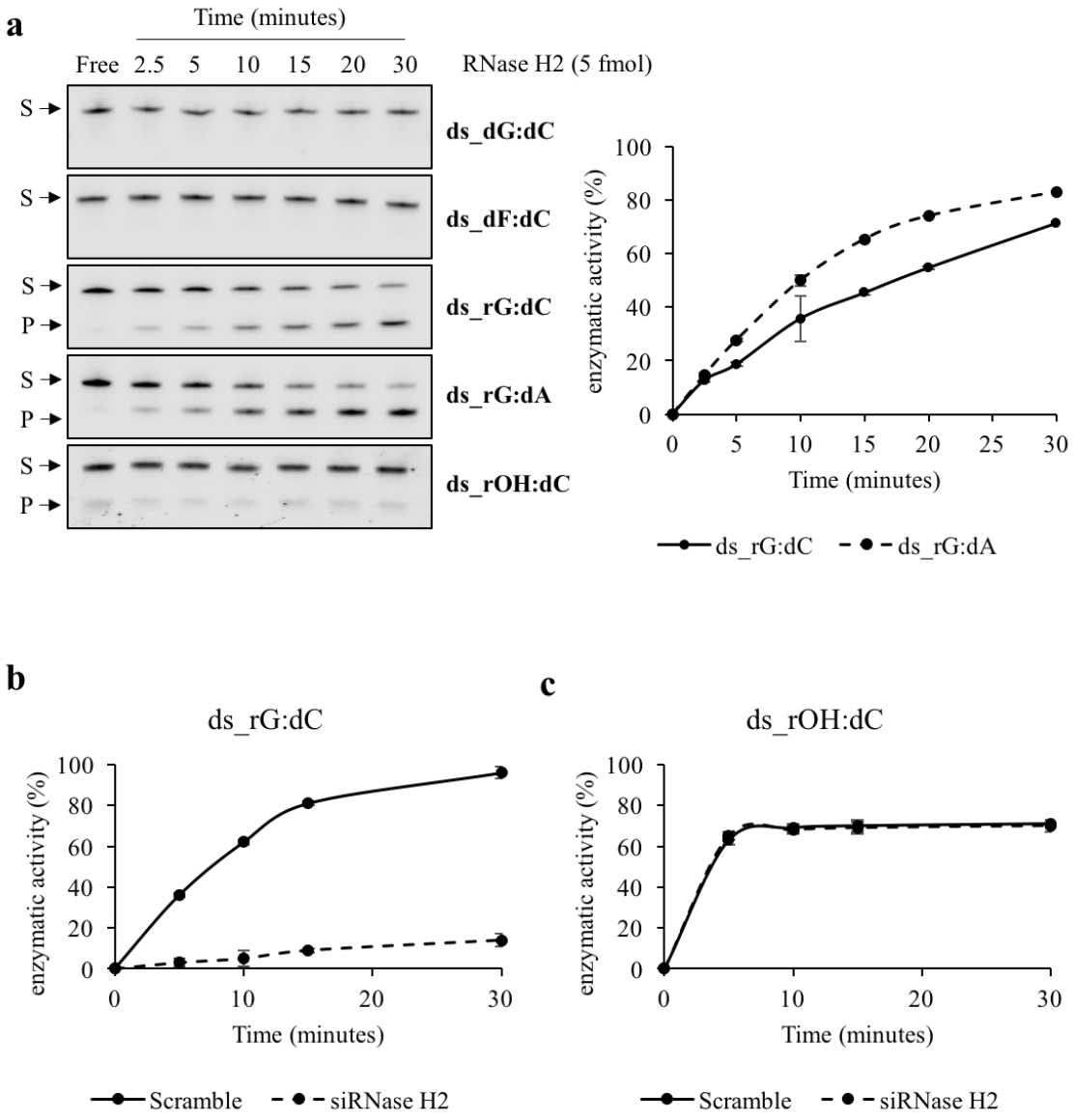
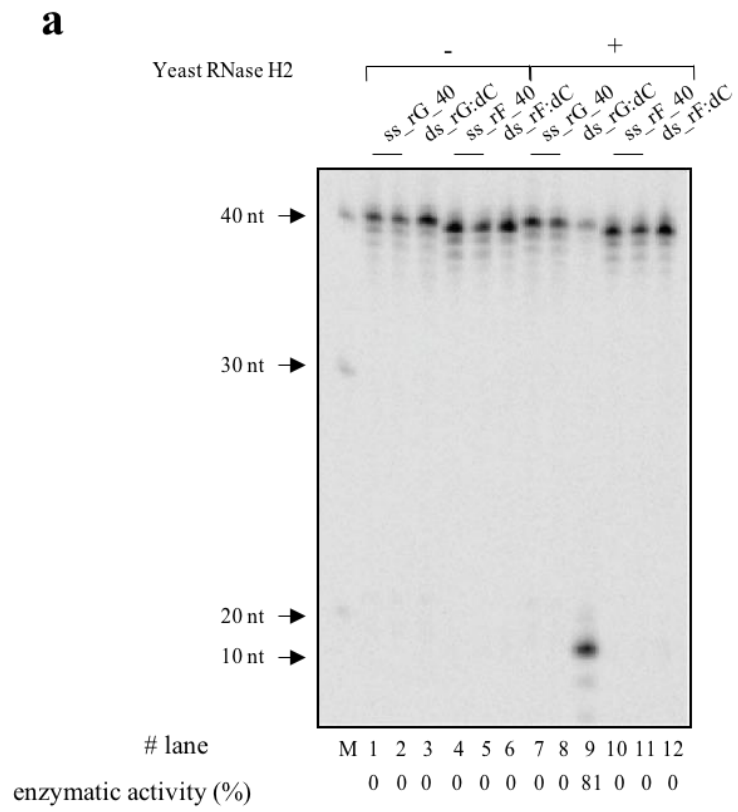


Figure 3.2 Human RNase H2 is not able to process an rAP site embedded in a duplex DNA substrate. (**a**) Representative denaturing polyacrylamide gel of oligonucleotides (25 nM) incision by recombinant human RNase H2 (0.5 nM). The reaction was performed in RNase H2-buffer (20 mM Tris-HCl, 25 mM KCl, 0.1% BSA, 0.01% Tween20, 4 mM $MgCl_2$, pH 7.4) for different time points, expressed in minutes and shown on the top of the figure, at 37°C. ds_dG:dC and ds_dF:dC oligonucleotides were used as negative controls whereas paired and mismatched ds_rG oligonucleotides as positive controls. S indicates the substrate position while P indicates the product position (*left*). Relative graph illustrating the time-course kinetics activity of the recombinant protein on ds_rG:dC and ds_rG:dA oligonucleotides. Data are expressed as mean ± SD of three independent technical replicas (*right).* (**b**) Graph illustrating the time-course kinetics activity of NCE on ds_rG:dC in control and RNase H2-knocked down conditions. Enzymatic reaction was

performed at 37°C in RNase H2-buffer with 500 ng of NCE. Data are expressed as mean ± SD of three independent technical replicas. Standard deviation values were always less than 10% of the mean of the experimental points. (**c**) Graph illustrating the time-course kinetics activity of NCE on ds_rOH:dC oligonucleotide in control and RNase H2-knocked down conditions. Enzymatic reaction was performed at 37°C in RNase H2-buffer with 500 ng of NCE. Data are expressed as mean ± SD of three independent technical replicas. Standard deviation values were always less than 10% of the mean of the experimental points.

To increase the stability of the abasic rNMP-containing oligonucleotide, we also used abasic substrates mimicked by tetrahydrofuran (F) residues, similarly to what commonly used for dNMP [133, 134]. We then tested whether RNase H2 from yeast or mouse may process an abasic rNMP, mimicked by a tetrahydrofuran ribonucleotide residue, embedded in a longer DNA sequence composed of 40-bp to evaluate a possible role of the length of the substrate in determining the inability of RNase H2 to process these substrates (**Figure 3.1b**). We used single-stranded (ss) or double-stranded (ds) DNA substrates containing an abasic rNMP site (rF), an rGMP (rG), an abasic dNMP (dF) or a dGMP as internal controls (**Table B.1**). As expected, *S. cerevisiae* and mouse RNase H2 cleaved at the single rG in a DNA duplex substrate (lane 9 in **Figure 3.3a,b**) and had no activity on the rG embedded in the ss substrate (lanes 7 and 8 in **Figure 3.3a,b**). Importantly, as we found for human RNase H2, also *S. cerevisiae* and mouse RNase H2 complexes were inactive on the abasic rNMP, ds_rF:dC (lane 12 in **Figure 3.3a,b**). Together these results demonstrate the inability of eukaryotic RNase H2, both from yeast and mammalian origins, to process an abasic rNMP incorporated in DNA independently from the nature of the abasic site (either 1'-OH or tetrahydrofuran residue) and the length of the substrate (either 25- or 40-mers).


a
Yeast RNase H2
-
+
ss_rG_40
ds_rG:dC
ss_rF_40
ds_rF:dC
ss_rG_40
ds_rG:dC
ss_rF_40
ds_rF:dC
40 nt
30 nt
20 nt
10 nt
lane
M 1 2 3 4 5 6 7 8 9 10 11 12
enzymatic activity (%)
0 0 0 0 0 0 0 0 81 0 0 0

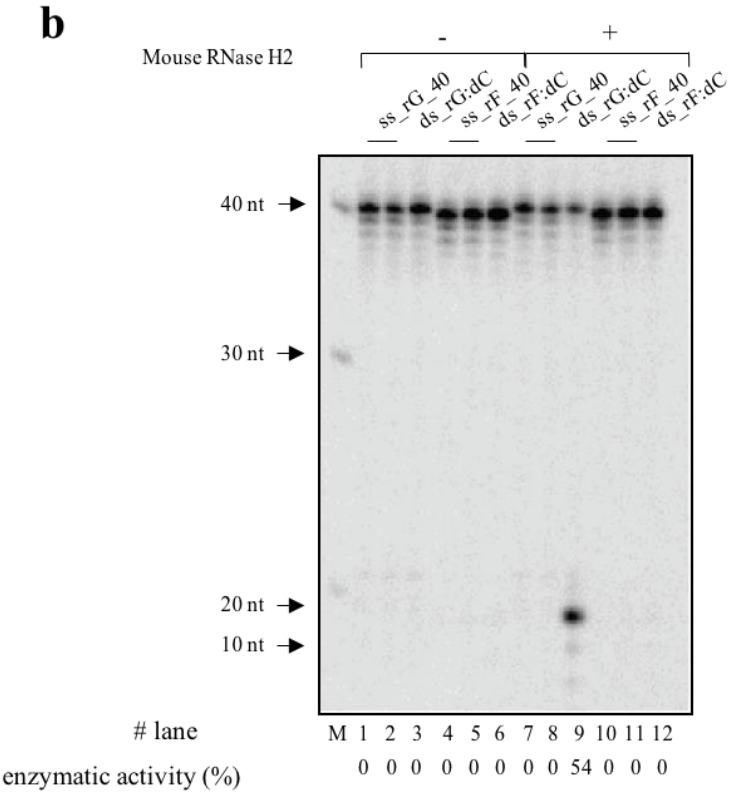

b
Mouse RNase H2
-
+
ss_rG_40
ds_rG:dC
ss_rF_40
ds_rF:dC
ss_rG_40
ds_rG:dC
ss_rF_40
ds_rF:dC
40 nt
30 nt
20 nt
10 nt
lane
M 1 2 3 4 5 6 7 8 9 10 11 12
enzymatic activity (%)
0 0 0 0 0 0 0 0 54 0 0 0

Figure 3.3 Mouse and yeast RNase H2 are not able to process an rAP site embedded in duplex DNA substrate. (a and **b)** Denaturing PAGE gels showing cleavage result using 10 nM of 40-mer radioactive substrate containing an rG or an rF site, without (lanes 1-6) or with (lanes 7-12) 10 nM of yeast RNase H2 protein (**a**) or 10 nM of mouse RNase H2 protein (**b**). All reactions were carried out at 37 °C for 1 hour in yeast/mouse reaction 1X buffer (15 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 5% glycerol, 1 mM DTT and 0.1 mg/mL BSA). M indicates the DNA ladder and the black arrows on the left of each panel show specific band sizes. Lanes 1,7 have ss-substrate containing rG (ss_rG_40), lanes 2,8 have ss-substrate containing ss_rG_40 that is cooled slowly at room temperature to demonstrate the absence of any self-annealing structures, and lanes 3,9 show the ds-substrate containing rG (ds_rG_dC). Lanes 4,10 have ss-substrate containing abasic ribo site (ss_rF_40), lanes 5,11 have ss_rF_40 that is cooled slowly in room temperature to observe any self-annealing; and lanes 6,12 have ds substrate containing abasic ribo site (ds_rF:dC). The percentages of cleavage of each reaction are displayed below the images as enzymatic activity (%).

3.4.2 Human APE1 is able to process an rAP site embedded in DNA through its endonuclease catalytic domain

To test the ability of APE1 to process abasic rNMPs in DNA, we purified human recombinant APE1 protein, as described in Materials and Methods section (**Figure B.4a**). The endonuclease activity of APE1 on ds_rOH:dC substrate was examined through cleavage assays. As a positive control for APE1 endonuclease activity, an oligonucleotide substrate containing a tetrahydrofuran residue mimicking the abasic site, called ds_dF:dC [133, 134], was used. As reported in **Figure 3.4a**, and measured through kinetics experiments in **Table 3.1**, APE1 processes the abasic rNMP as efficiently as the canonical deoxy-abasic site having a lower affinity for the ds_rOH:dC than the ds_dF:dC (11-fold increase of the K_M) but a higher catalytic rate (27-fold increase in the k_{cat}/K_M ratio) (**Table 3.1)**. Moreover, APE1 was unable to process the rG- containing oligonucleotide, which is the preferential substrate of RNase H2 enzyme. To further characterize the enzymatic

activity of APE1, we used the purified recombinant mutant APE1 E96A protein, in which the missense mutation of the residue in the catalytic site, characterized by the substitution of the glutamic acid in position 96 with alanine, causes a decreased enzymatic activity of the protein, lacking the ability to coordinate the Mg^{2+} ion in the catalytic site [133, 135, 136] (**Figure B.4b**). In addition, we used the purified recombinant mutant APE1 NΔ33 protein, in which the first 33 N-terminal residues, responsible for RNA-protein interaction but not affecting its enzymatic activity, have been deleted [5, 66] (**Figure 3.4b**). Following incubation of ds_rOH:dC with APE1 E96A mutant, there was barely any endonuclease activity, whereas the activity of APE1 NΔ33 mutant was comparable with that of the APE1 WT protein (**Figure 3.4b,c** and **Figure B.4c**). These data demonstrate that the catalytic domain of APE1 is responsible for recognizing and cleaving a rAP site in dsDNA and that the N-terminal domain does not play any major role in the enzymatic activity on this substrate and that AP endonucleolytic activity on abasic rNMPs is intrinsic to the purified protein.

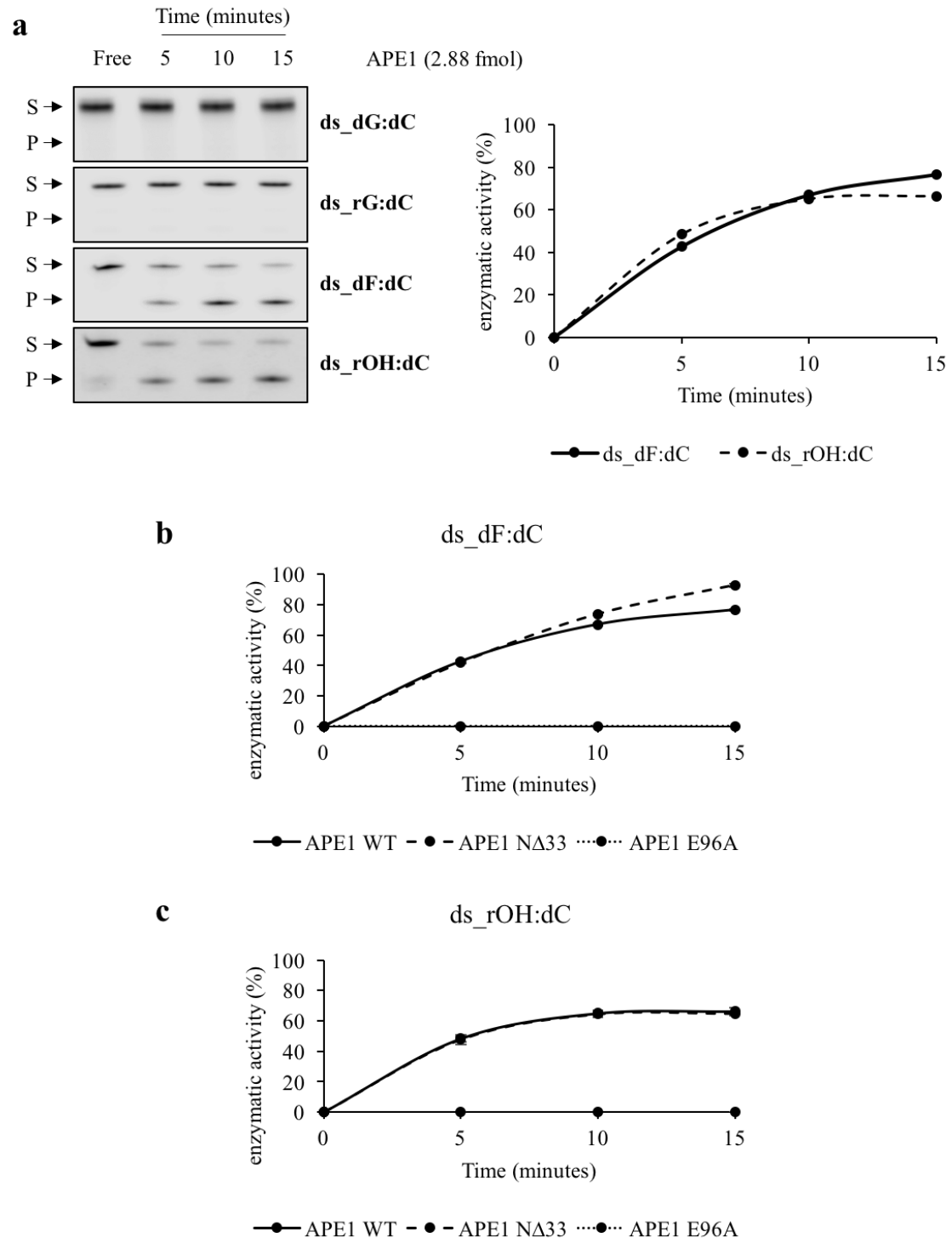


Figure 3.4 Human PE1 efficiently processes an rAP site embedded in a duplex DNA substrate. (**a**) Representative denaturing polyacrylamide gel of oligonucleotides (0.25 µM) incision by recombinant human APE1 (0.288 nM). The reaction was performed in APE1-buffer (20 mM Tris-HCl, 100 mM KCl, 0.1% BSA, 0.01% Tween20, pH 7.4) for

different time points, expressed in minutes and shown on the top of the figure, at 37°C. ds_dG:dC and ds_rG:dC oligonucleotides were used as negative controls, whereas ds_dF:dC oligonucleotide as positive control. S indicates the substrate position, while P indicates the product position (left). Relative graph illustrating the time-course kinetics activity of the recombinant protein on ds_dF:dC and ds_rOH:dC oligonucleotides. Data are expressed as mean ± SD of three independent technical replicas. Standard deviation values were always less than 10% of the mean of the experimental points (right). (**b**) Graph illustrating the time-course kinetics activity of APE1 mutants on ds_dF:dC oligonucleotide. Data are expressed as mean ± SD of three independent technical replicas. Standard deviation values were always less than 10% of the mean of the experimental points. (**c**) Graph illustrating the time-course kinetics activity of APE1 mutants on ds_rOH:dC oligonucleotide. Data are expressed as mean ± SD of three independent technical replicas. Standard deviation values were always less than 10% of the mean of the experimental points.

Table 3.1 Kinetic parameters for APE1 endonuclease activity on different substrates

[APE1] ($\times 10^{-3}$ nM)	Substrate	K_M (nM)	V_{MAX} (nM/min)	k_{CAT} (min^{-1})	k_{CAT}/K_M $(min*nM)^{-1}$
75	ds_dF:dC	14.2±6.98	0.95±0.33	12.7±4.39	0.95±0.16
3.125	ds_rOH:dC	158±41.79	12.66±3.71	4054±1191.8	26±1.96

Kinetic parameters (K_M, V_{MAX} and k_{CAT}) were calculated from the measurements of the endonucleolytic reaction rates for APE1 on ds_dF:dS and ds_rOH:dS substrates. As described by Fantini *et al* [130], increasing concentrations of the substrate were incubated with a selecting concerntration of the protein (see first column in **Table 3.1**) in a time-course experiment. Kinetic values were calculated using a Lineweaver-Burk plot analysis and represent the mean ± SD of three independent experiments.

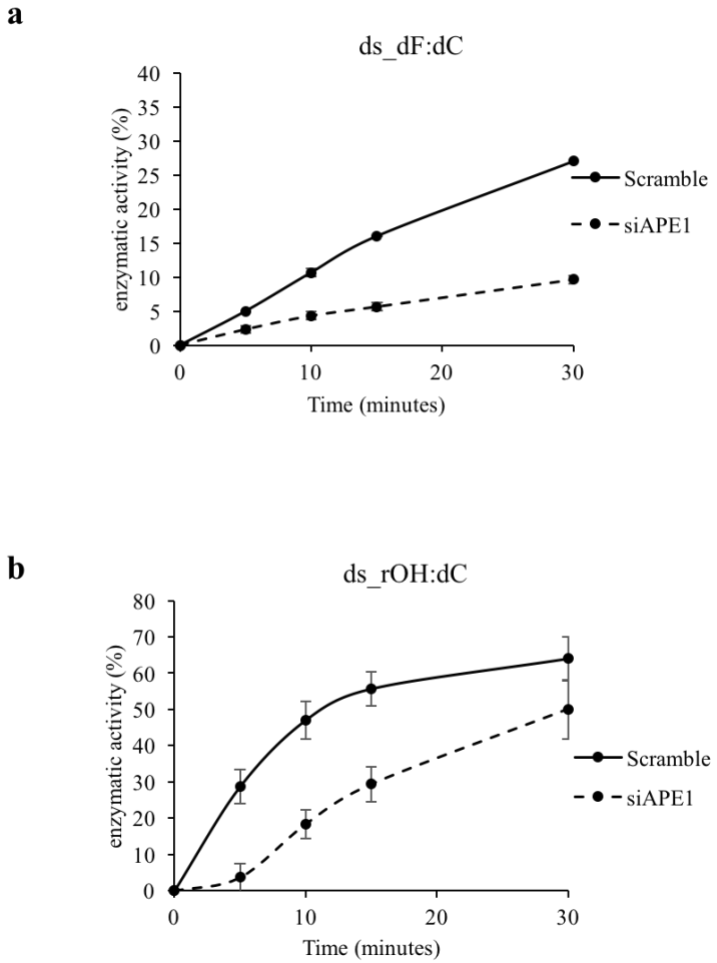


Figure 3.5 APE1 knock down in human cells impairs the processing of an rAP site embedded in a duplex DNA substrate. (**a**) Graph illustrating the time-course kinetics activity of NCE on ds_dF:dC in control and APE1-knocked down conditions. Data are expressed as mean ± SD of three independent technical replicas. Standard deviation values were always less than 10% of the mean of the experimental points. (**b**) Graph illustrating the time-course kinetics activity of NCE on ds_rOH:dC oligonucleotide in control and APE1-knocked down conditions. Enzymatic reaction was performed at 37°C in APE1-buffer with 10 ng of NCE. Data are expressed as mean ± SD of three independent technical replicas.

To confirm that APE1 is the major enzyme capable of cleaving ribo-abasic containing sites in cells, we used nuclear cell extracts in which APE1 was knocked-down through specific siRNAs (**Figure B.3a**), as explained in Material and Methods. The endonuclease activity

of APE1-kd (siAPE1) cell extracts, was reduced on both ds_dF:dC (**Figure 3.5a** and **Figure B.4d**) and ds_rOH:dC (**Figure 3.5b** and **Figure B.4d)**, as compared to its respective control SCR-treated extracts. As Western blot analysis shows (**Figure B.3a**), the expression of APE1 protein did not exert any effect on the expression of RNase H2 protein itself, demonstrating that the observed reduction of the processing activity of the abasic rNMP-containing substrate, observed with APE1-kd cell extracts, was likely due to the reduced expression of the APE1 protein. We further examined the specificity of the enzymatic activity of APE1 using the tetrahydrofuran ribonucleotide mimicking an abasic residue.

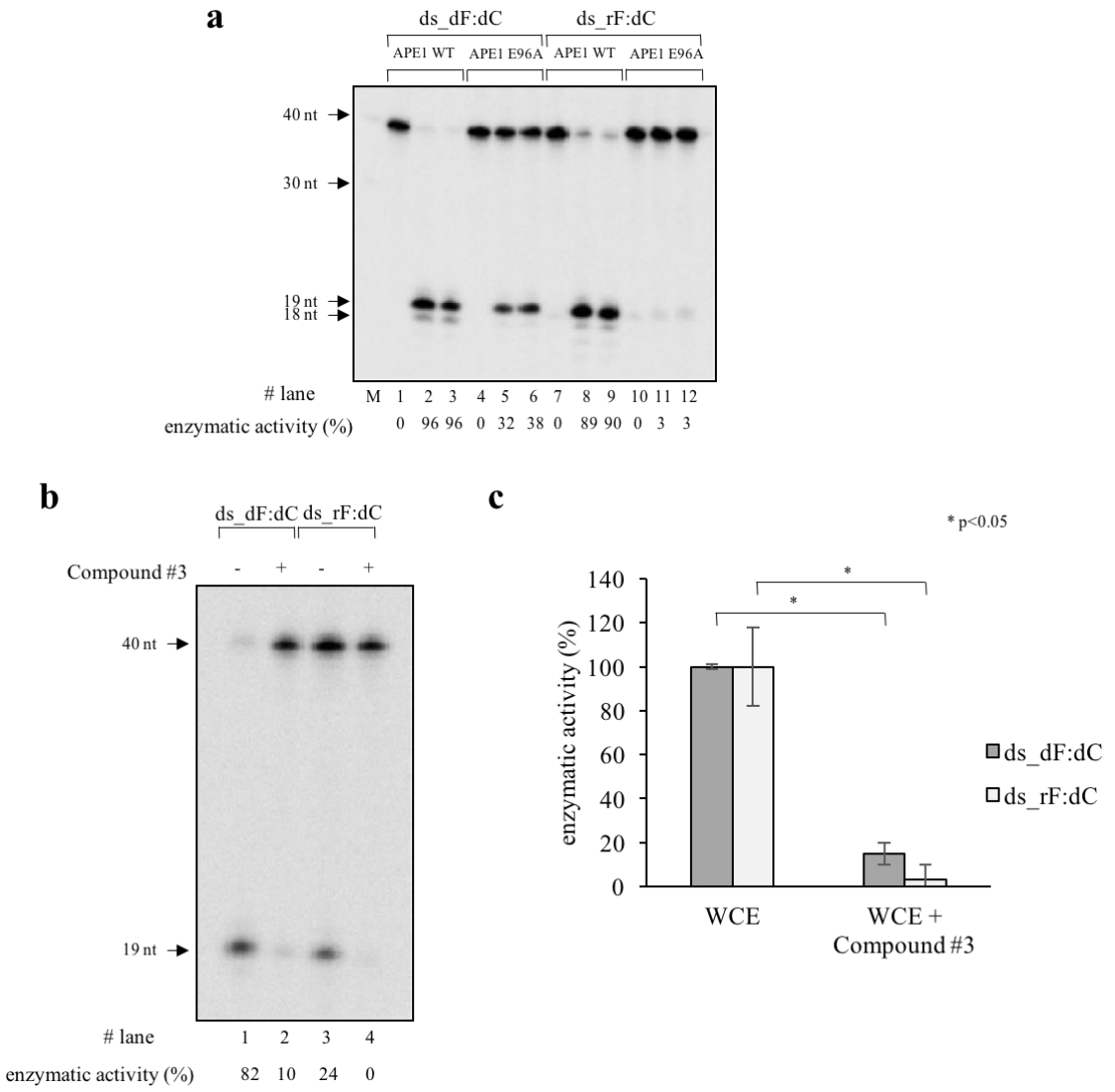


Figure 3.6 Recombinant human APE1 efficiently processes a tetrahydrofuran ribonucletotide mimicking an abasic residue site (rF) embedded in a duplex DNA substrate. (**a**) Cleavage result of 10 nM of ^{32}P double-stranded oligonucleotides ds_dF:dC or ds_rF:dC substrates with different concentrations of APE1 WT protein (lanes 1-3 and 7-9) and mutant APE1 E96A protein (lanes 4-6 and 10-12) at 37 °C for 1 hour in APE1-reaction 1X buffer containing 50 mM Tris-HCl pH 7.5, 50 mM KCl, 10 mM $MgCl_2$, 0.001 mg/mL BSA and 0.05% Triton X-100. First lane on the left, M is a ssDNA ladder (barely visible) and the black arrows on the left show specific band sizes. Both oligonucleotides were incubated with 0 nM (lanes 1,4,7,10), 5 nM (lanes 2,5,8,11) or 10 nM (lanes 3,6,9,12) of APE1 WT protein (lanes 1-3 and 7-9) or mutant E96A protein (lanes 4-6 and 10-12), respectively. The percentages of cleavage of each reaction are displayed below the image as enzymatic activity (%). (**b**) Cleavage result of 10 nM ds_dF:dC or ds_rF:dC radioactive substrates using 12.5 ng of whole HeLa cell extracts untreated (lanes 1,3) and treated (lanes 2,4) with 0.20 nM of Compound #3, a specific APE1-endonuclease inhibitor, at 37 °C for

10 min. The black arrows on the left of the gel image indicate the size of uncut and cut substrates following denaturation. The cleavage percentage of this experiment is displayed below the image as enzymatic activity (%). (**c**) Histograms showing data from four independent experimental replicas shown in panel (**b**) with ranges as bars. *P* values of less than 0.05 are marked by asterisk. WCE, whole cell extracts.

We tested the activity of the APE1 E96A mutant to cleave at a ds_rF:dC substrate compared to that of APE1 WT (**Figure 3.6a**). As a control, the activity on the ds_dF:dC substrate was also analyzed. As it can be observed, mutant E96A showed a reduced cleavage on the ds_rF:dC substrate. Indeed, APE1 WT gave a 96% of cleavage at 5 and 10 nM (lanes 2 and 3), whereas mutant APE1 E96A had 32% cleavage at 5 nM and 38% at 10 nM (lanes 5 and 6). Cleavage of ds_rF:dC by APE1 WT was 89% at 5 nM and 90% at 10 nM (lanes 8 and 9), and there was minimal activity for APE1 E96A on the ds_rF:dC substrate in the same conditions (lanes 11 and 12). Moreover, pre-treatment of cell extracts with 0.2 nM of Compound #3 (i.e. N-(3-(benzo[d]thiazol-2-yl)-6-isopropyl-4,5,6,7-tetrahydrothieno[2,3-c]pyridin-2-yl)acetamide)), a specific APE1 endonuclease inhibitor [134, 137], exerted a significant inhibitory effect upon APE1 enzymatic activity on both ds_dF:dC in DNA and ds_rF:dC substrates demonstrating that the main enzymatic activity of cell extracts was due to APE1 function (**Figure 3.6b,c**).These data demonstrate that human APE1 is the major enzyme capable of specifically cleaving at abasic rNMPs in DNA while being unable to process normal rNMPs, which are the preferential substrates of RNase H2.

3.4.3 Human RNase H2 does not process an r8oxoG embedded in a duplex DNA

An abasic site can be generated spontaneously or following the processing of an oxidized lesion from a specific glycosylase. While known glycosylases (such as OGG1) are responsible for this activity on 8oxo-dG, no enzyme is known to be able to process the oxidized rG substrate. We then focused our attention on r8oxoG removal. The r8oxoG containing oligonucleotide was in-house synthesized and, as observed from MALDI-MS analysis and HPLC purification (**Figure B.5a,b)**, the undesirable presence of secondary products of the chemical synthesis of this substrate can be excluded.

First, we investigated whether human RNase H2 was able to recognize and cleave at r8oxoG site using an oligonucleotide containing this type of lesion called ds_r8oxoG:dC (**Figure 3.7a,** see also **Table B.1, Figure B.1**). In these experiments, we always compared the specific enzymatic activity of RNase H2 with that exerted on the canonical ds_rG:dC substrate, as positive control, and using the ds_d8oxoG:dC oligonucleotide as negative control. As **Figure 3.7b** shows, while the ds_rG:dC oligonucleotide was efficiently processed by RNase H2, the same activity was not observed for ds_r8oxoG:dC. As expected, the d8oxoG-containing oligonucleotide was not cleaved by RNase H2. These data confirm that RNase H2 was not able to process modified rNMPs embedded in DNA.

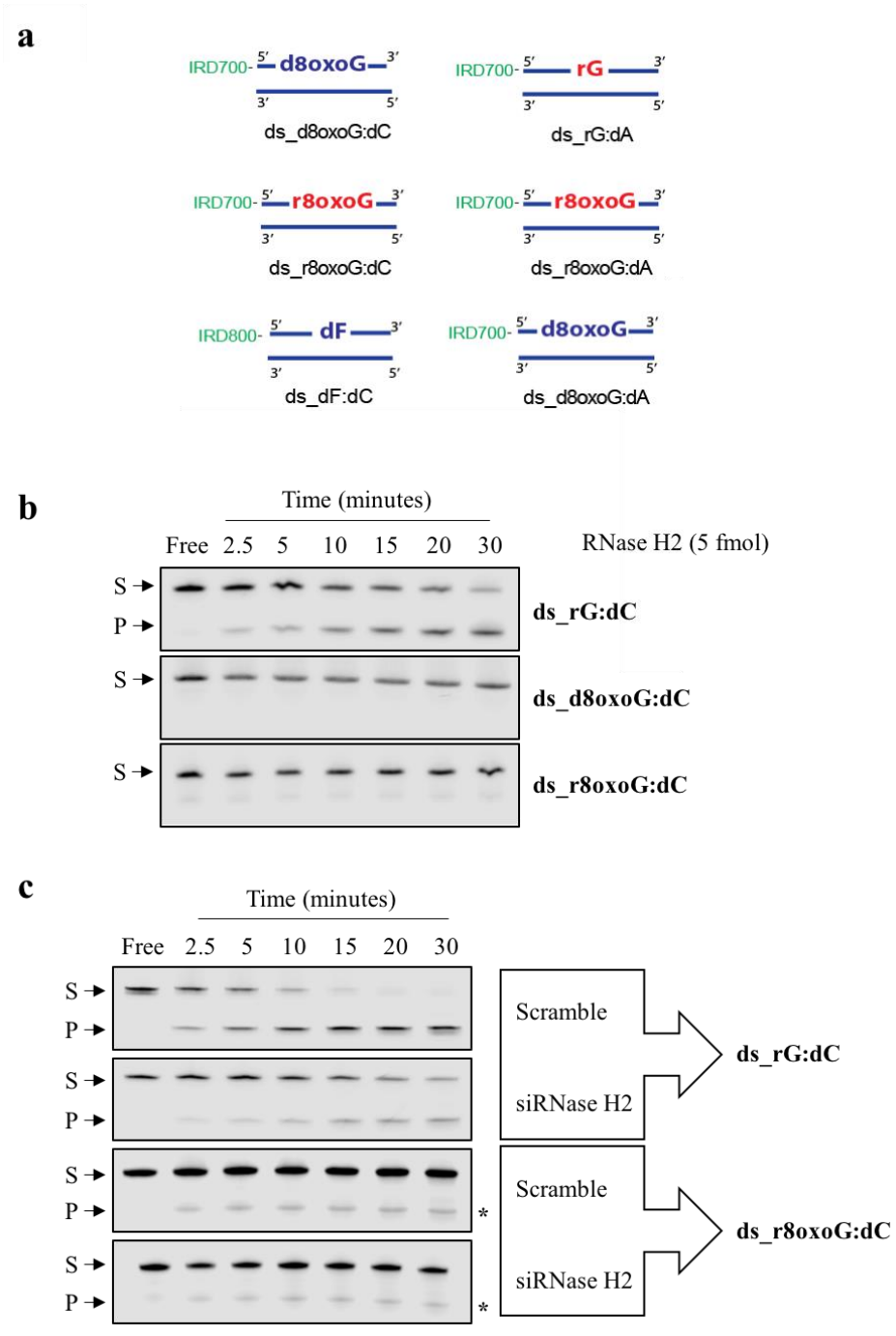


Figure 3.7 Human RNase H2 is not able to process an r8oxoG site embedded in a duplex DNA substrate. (**a**) Scheme of ds_DNA 25-mer substrates (single strand (ss) oligonucleotide annealed to ss_dC containing- or ss_dA containing-complementary oligonucleotide) containing a d8oxoG, r8oxoG and dF site in the 13th position. The 'IRD700' and 'IRD800' in green indicates IRDye 800 phosphoramidite and IRDye700 phosphoramidite dye tagged at the 5' end. (**b**) Representative denaturing polyacrylamide gel of oligonucleotides (25 nM) incision by human recombinant RNase H2 (5 fmol).

Reaction was performed in RNase H2-buffer (20 mM Tris-HCl, 25 mM KCl, 0.1% BSA, 0.01% Tween20, 4 mM $MgCl_2$, pH 7.4) at 37°C. ds_rG:dC and ds_d8oxoG:dC oligonucleotides were used as positive and negative controls, respectively. Time points are shown on the top of the figure. S indicates the substrate position, while P indicates the product position. (**c**) Representative denaturing polyacrylamide gel of oligonucleotides (0.25 µM) incision by nuclear HeLa cell extracts (NCE). In order to discriminate the activity of RNase H2, 500 ng of NCE in which RNase H2 expression was previously knocked down through specific siRNA (indicated as siRNase H2) were tested in comparison to control cells (Scramble) at different time points (minutes), shown on top of the figure. The endoribonuclease activity detected for r8oxoG oligonucleotide was indicated with an asterisk on the right side of each panel. S indicates the substrate position while P indicates the product position.

Similarly, we confirmed these data using RNase H2-kd nuclear extracts from HeLa cells. As expected, control nuclear extracts (Scramble) displayed a time-dependent endoribonuclease activity on ds_rG:dC, whereas the down regulation of RNase H2 protein expression (siRNase H2) was associated with a marked reduction of the endoribonuclease activity on the same substrate (**Figure 3.7c**). On the contrary, once we tested the ability of the nuclear extracts on ds_r8oxoG:dC oligonucleotide, we found only a weak endoribonuclease activity on it (see the band indicated by an asterisk), which was not affected by RNase H2 silencing (**Figure 3.7c**). These data support the conclusion that the r8oxoG site in DNA is not recognized by human RNase H2.

3.4.4 OGG1 has neither lyase nor glycosylase activities on oxidized rG substrate

We then tested whether enzymes of the BER pathway may be involved in processing the r8oxoG substrate. To this purpose, we used recombinant purified human OGG1 and APE1 proteins (**Figures B.4a** and **B.5c**). OGG1 protein belongs to the bi-functional glycosylases family having both lyase and glycosylase activities on oxidized dG [138]. We examined the processing activity of OGG1 on r8oxoG substrate in comparison to the d8oxoG-

containing oligonucleotide, as a positive control. First, we tested the ability of OGG1 to recognize an r8oxoG site through electrophoretic mobility shift assay (EMSA). As shown in **Figure 3.8a**, increasing amount of recombinant OGG1 formed a stable retarded complex with the r8oxoG oligonucleotide in a dose-dependent manner. As confirmed in the data shown in **Figure B.5e**, OGG1 binding was specific for the modified r8oxoG- containing oligonucleotide (lanes 2 and 3). Indeed, the same retarded band was observed when the recombinant OGG1 was incubated with the positive control ds_d8oxoG:dC (lanes 11 and 12) but not with the negative controls: ds_rG:dC (lanes 5 and 6) and ds_dG:dC (lanes 8 and 9).

Furthermore, we tested the lyase activity of OGG1 on different substrates. **Figure 3.8b** shows that when we incubated increasing amounts of recombinant protein for 30 minutes at 37°C with different substrates, OGG1 was able to process only the canonical substrate ds_d8oxoG:dC in a dose response manner (**Supplementary Figure B.5d**), whereas no lyase activity was apparent for any of the other substrates used, including the ds_r8oxoG:dC and the ds_r8oxoG:dA.

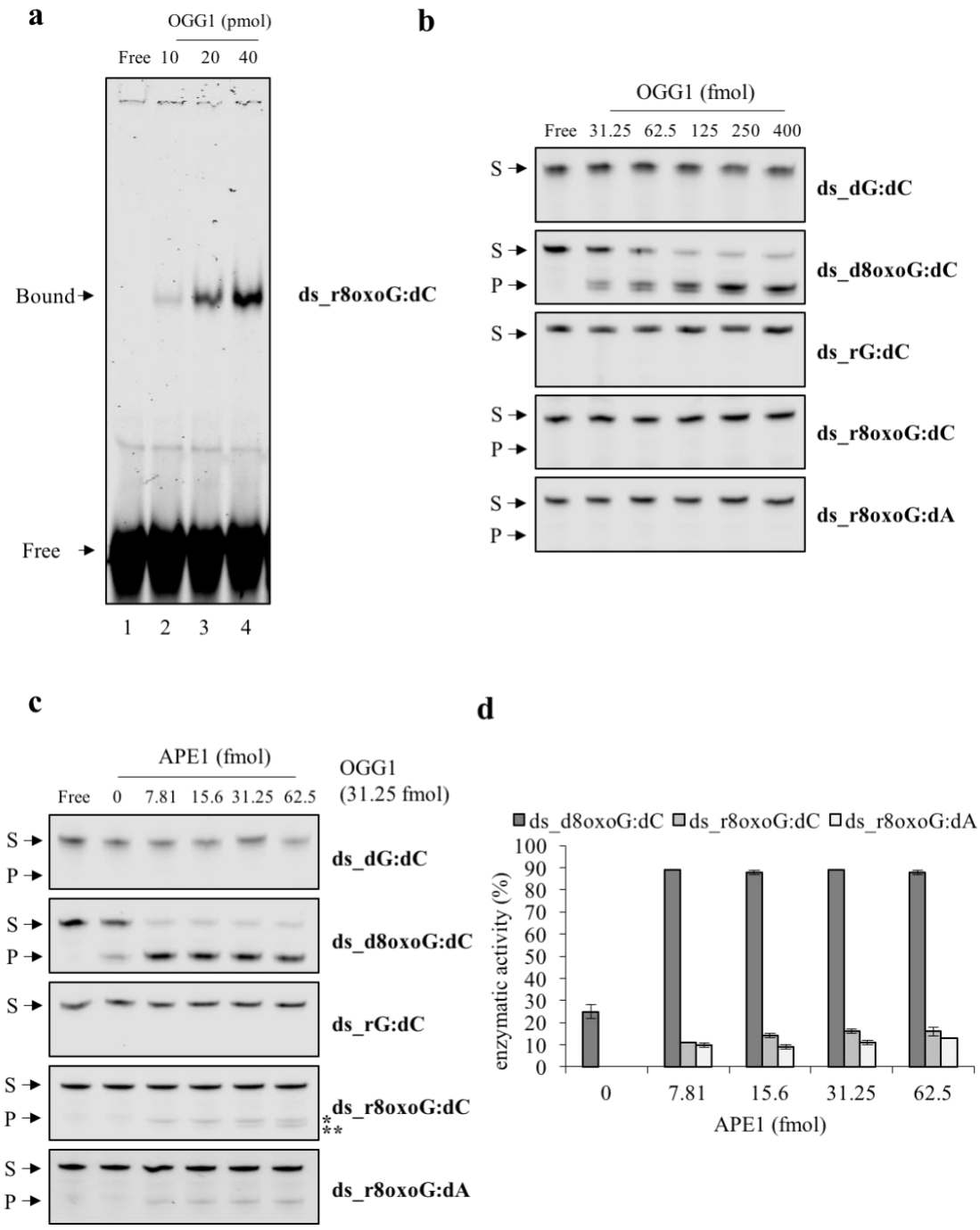


Figure 3.8 Human OGGI has neither lyase nor glycosylase activities on r8oxoG-containing oligonucleotide. (**a**) Representative native EMSA polyacrylamide gel of OGG1 binding on ds_r8oxoG:dC oligonucleotide (25 nM) is shown. The 'Bound' arrow indicates the retarded complex between OGG1 and the probe whereas the 'Free' arrow the unbound substrate. Amounts of OGG1 protein, expressed in *pico* moles, are shown on the top of the figure. Reactions were performed as explained in "Material and Methods" section. (**b**) Representative denaturing polyacrylamide gel of lyase activity of human

OGG1 on different duplex DNA oligonucleotides (25 nM). Doses of OGG1 protein expressed in *femto* moles are shown on the top of the figure. Reactions were performed in OGG1-buffer (20 mM Tris-HCl, 100 mM KCl, 0.1% BSA, 0.01% Tween20, pH 7.4) at 37°C for 30 minutes. S indicates the substrate position, while P indicates the product position. (**c**) Representative denaturing polyacrylamide gel of incision by different doses of APE1 co-incubated with a fixed amount of OGG1 (3.125 nM) on different duplex DNA oligonucleotides (25 nM) in order to investigate glycosylase activity of OGG1. Different doses of APE1 protein, expressed in *femto* moles, are shown on top of the figure. Reactions were performed in a buffer containing 20 mM Tris-HCl, 100 mM KCl, 0.1% BSA, 0.01% Tween20, pH 7.4 at 37°C for 30 minutes. S indicates the substrate position while P indicates the product position. Moreover, at the right of the panel, a longer product of about 12 nucleotides is indicated by an asterisk whereas a smaller one of 11 nucleotides is indicated by a double asterisk. (**d**) Histograms represent the dose response of OGG1 glycosylase activity on ds_d8oxoG:dC and paired and mismatched ds_r8oxoG oligonucleotides. ds_d8oxoG oligonucleotide was used as a positive control whereas ds_dG:dC and ds_rG:dC oligonucleotides were used as negative controls. The activity is reported as percentage of substrate converted to product. Data are expressed as mean ± SD of three independent technical replicas.

Since OGG1 is the major glycosylase enzyme in the BER pathway, coordinating with the downstream endonuclease APE1, which is able to recognize and process the abasic site generated by the glycosylase activity of OGG1 [65, 139, 140], we measured its glycosylase activity on the same substrates using recombinant purified APE1 (**Figure 3.8c**). In this case, we co-incubated a fixed amount of recombinant OGG1 with increasing amounts of APE1 for 30 minutes. While OGG1 displayed a robust glycosylase activity on the canonical ds_d8oxoG:dC substrate, particularly in the presence of the APE1 protein, we detected only a weak activity on ds_r8oxoG:dC oligonucleotide (indicated by a single asterisk) (**Figure 3.8d**). Moreover, the presence of an additional higher mobility band, increasing as a function of APE1 concentration (indicated with a double asterisk) was observed only in the case of the ds_r8oxoG:dC substrate. We conclude that OGG1 has neither lyase nor glycosylase activity on the r8oxoG substrate, and that APE1 can weakly

process this substrate alone. With respect to APE1 activity, a detailed description is explained in the next paragraph.

3.4.5 APE1 has a weak endo-/exo-nuclease activities on the r8oxoG-containing substrate depending on Mg^{2+} concentration and on the presence of its N-terminal domain

Based on the above presented data, we then checked whether APE1 '*per se*' had any endoribonuclease activity on ds_r8oxoG:dC substrate (**Figures 3.9 and 3.10**). Compared to the ds_dF:dC substrate, APE1 displayed a modest, though significant, processing activity on both ds_r8oxoG:dC and ds_r8oxoG:dA oligonucleotides, while no activity was observed in the case of the dG- and the d8oxoG-containing substrates (**Figure 3.9a,b**), as expected. As observed above, the appearance of an additional faster migrating cleavage product (indicated by a double asterisk corresponding to a 11-nt product in **Figure 3.8d**) was visible in the case of the ds_r8oxoG:dC substrate, which might be associated with a recently identified 3'-exonuclease activity by the protein [124]. We checked the occurrence of the cleavage at the expected ribonucleotide sites, by oligonucleotide sequences of increasing length, ranging from 10 to 16 nucleotides, as molecular markers (**Figure B.6a**) and through alkaline hydrolysis experiments (**Figure B.6b**). As it is visible (**Figure B.6a**), cleavage products of the ds_r8oxoG:dC oligonucleotide were of the expected size and comprised between 11- and 12- nucleotides and are thus compatible with endonucleolytic cleavage occurring only at the 5' side of the lesion (fragment 12-nt long) and with a 3'-exonuclease activity giving rise to the fragment of 11-nt long. In order to exclude that the observed cleavage product was due to the processing of a residual non-annealed oligonucleotide possibly present after the annealing reaction, we incubated APE1 protein with single stranded oligonucleotide (ss_r8oxoG) and compared the cleavage product with

the annealed oligonucleotide (ds_r8oxoG:dC). Comparing the result with both ss_dF and ds_dF:dC oligonucleotides, a product was detectable only using the double stranded oligonucleotides as substrates. No bands were observed using the ss_r8oxoG oligonucleotide (**Figure B.6c**) demonstrating the requirement for secondary structured oligonucleotide sequences for efficient enzymatic activities by APE1. In contrast, using the oligonucleotide containing the mismatched ds_r8oxoG:dA, most of the fragments produced after incision by the AP endonucleolytic activity (indicated with a single asterisk) were not further degraded by the exonucleolytic activity (**Figure 3.9a**). Therefore, these data demonstrate that the APE1 enzymatic activity on the r8oxoG substrate requires a dsDNA molecule and exonuclease activity is dependent on the paired nucleotide, possibly as a consequence of a different stereochemical geometry between the 8oxoG:A and the 8oxoG:C.

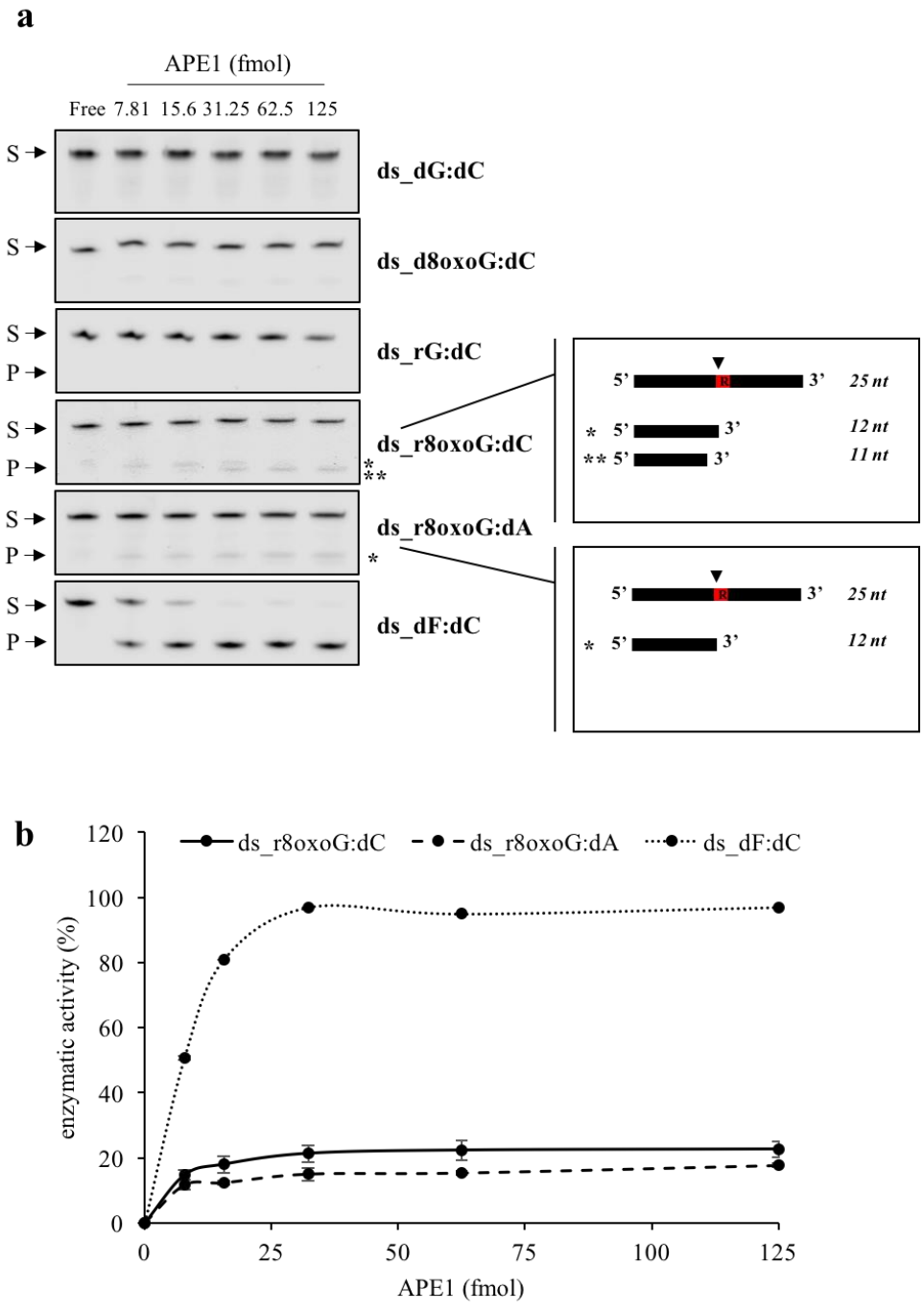


Figure 3.9 Human APE1 shows a weak endo- and a 3'-exonuclease activities on r8oxoG substrate. (**a**) Representative denaturing polyacrylamide gel of APE1 incision on different duplex DNA oligonucleotides (25 nM), in which ds_dF:dC oligonucleotide was used as a positive control, whereas ds_dG:dC and ds_rG:dC oligonucleotides were used as negative controls. The doses of APE1 protein used, expressed in *femto* moles, are shown on the top of the figure. On the right side, a schematic representation of the cleavage

products, showing the position of the ribonucleotide (red box with R) embedded in the DNA oligonucleotide and the APE1 cleavage on it, producing a longer product of about 12 nucleotides (*) and a smaller one of 11 nucleotides (**). Reactions were performed in APE1-buffer containing 20 mM Tris-HCl, 100 mM KCl, 0.1% BSA, 0.01% Tween20, pH 7.4 for 30 minutes at 37°C. S indicates the substrate position while P indicates the product position. (**b**) Relative graph indicates a dose-response APE1 activity on paired and mismatched ds_r8oxoG:dC oligonucleotide in comparison to ds_dF:dC positive control. Data are expressed as mean ± SD of three independent technical replicas.

It has been previously demonstrated that the exonuclease activity of APE1 strictly depends on salt concentrations [141]. We therefore tested whether the 3'-exonuclease activity observed on the ds_r8oxodG:dC shared some common features (in terms of dependence on the ionic strength conditions) with the 3'-exonuclease activity on mispaired DNA, as previously described [141]. Firstly, we determined the optimal $MgCl_2$ (**Figure 4.10a**) and KCl (**Figure 4.10b**) concentrations required for the 3'-exonuclease activity. Indeed, the 3'-exonuclease activity was present up to a concentration of 2 mM $MgCl_2$. An inhibitory effect was apparent at $MgCl_2$ concentrations above 4 mM. At the same time, the 3'-exonuclease activity was poorly affected at KCl concentration equal to 100 mM. These results are in line with previous data on 3'-mispaired DNA [141] and suggest that the observed 3'-exonuclease activity strongly depends on the electrostatic interaction of APE1 with the substrate during the cleavage reaction and with the role of Mg^{2+} ions.

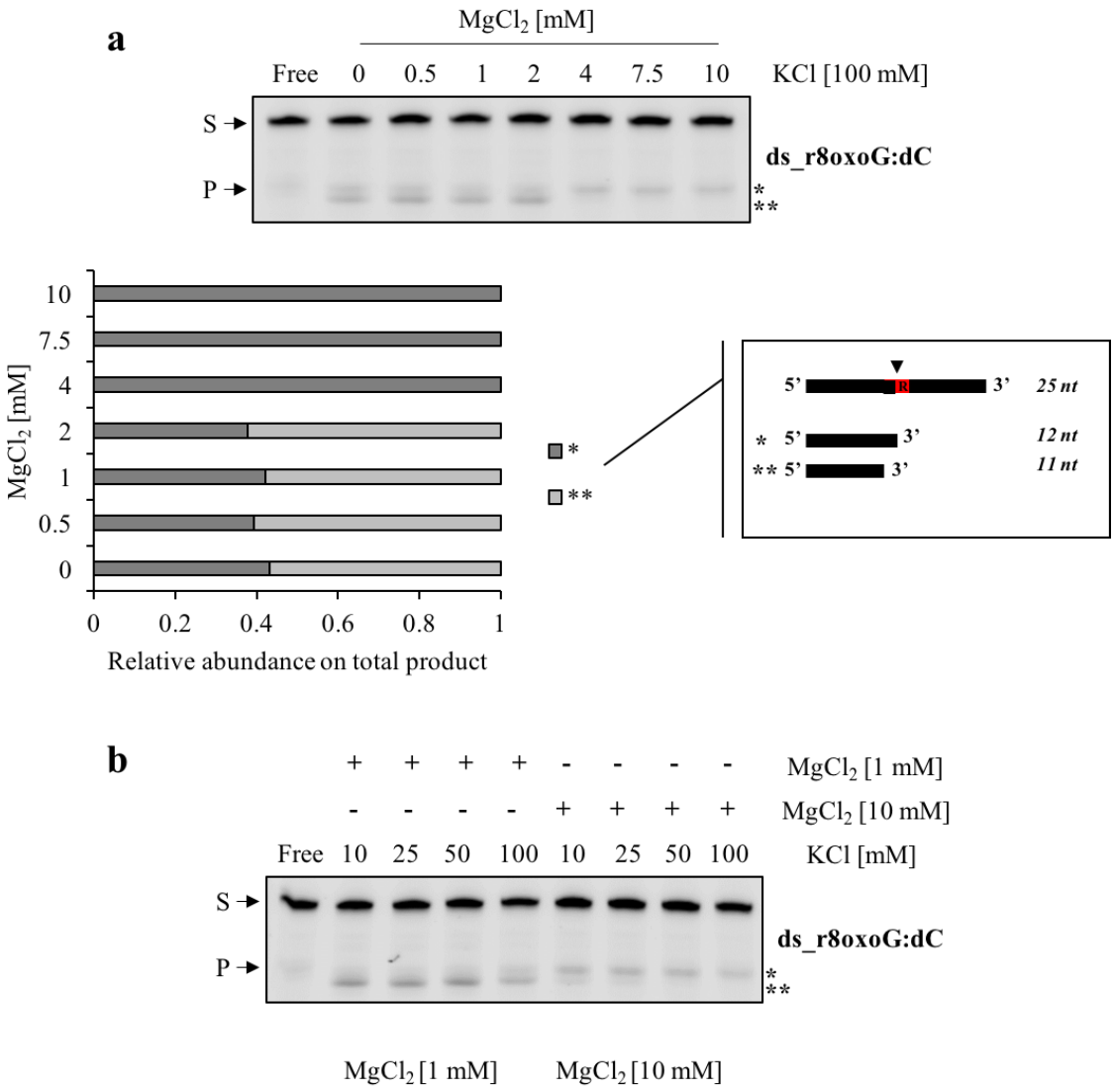


Figure 3.10 Human APE1 activities on r8oxoG substrate depend on mono- and di-valent cations. (**a**) Representative denaturing polyacrylamide gel of APE1 (5 nM) incision on ds_r8oxoG:dC oligonucleotide under different $MgCl_2$ concentrations that is expressed in mM (top) performing a reaction long 30 minutes in APE1 buffer. S indicates the substrate position while P indicates the product position. Moreover, at the right of the panel, a longer product of about 12 nucleotides is indicated by an asterisk whereas a smaller one of 11 nucleotides is indicated by a double asterisk Also shown on the right is a schematic representation of the cleavage products, showing the position of the ribonucleotide (red box with R) embedded in the DNA oligonucleotide and the APE1 cleavage on it, producing a longer product of about 12 nucleotides (*) and a smaller one of 11 nucleotides (**). Relative graph shows the ratio between two products obtained as a function of $MgCl_2$ concentration (bottom). (**b**) Representative denaturing polyacrylamide gel of APE1 (5 nM) incision on ds_r8oxoG:dC oligonucleotide under different KCl concentrations in combination with two different $MgCl_2$ concentrations, 1 mM and 10 mM (top) performing

a reaction long 30 minutes in a buffer containing 20 mM Tris-HCl, 0.1% BSA, 0.01% Tween20, pH 7.4. S indicates the substrate position, while P indicates the product position. Moreover, at the right of the panel, a longer product of about 12 nucleotides is indicated by an asterisk whereas a smaller one of 11 nucleotides is indicated by a double asterisk. Relative graph shows product levels in association with different salts concentrations (bottom).

After choosing the optimal salts conditions, in which both endo- and exo- activities of APE1 (100 mM KCl and 1 mM $MgCl_2$) are present, we then evaluated whether the enzymatic activity of APE1 on ds_r8oxoG:dC was dependent on the same catalytic site responsible for the endonuclease activity observed on abasic dsDNA and abasic rNMP. To this aim, the enzymatic activity of the E96A mutant was compared to that of the WT protein (**Figure B.6d**). These data demonstrate that the APE1 E96A mutant has a reduced endoribonuclease activity showing no 3'-exonuclease activity over the ds_r8oxoG:dC substrate. Due to the effect of salt concentration on this latter activity, we also tested the enzymatic activity of the APE1 NΔ33 deletion mutant (**Figure B.6d**). Interestingly, while this protein retained the endoribonuclease activity of the WT protein, its 3'-exonuclease activity was abolished. Moreover, treatment with APE1 inhibitor Compound #3 confirmed that the catalytic site responsible for the endonuclease activity is also responsible for the endoribonuclease activity over the r8oxoG substrate (**Figure B.6e**). These results show that APE1 has a weak, though significant, endoribonuclease activity on r8oxoG substrates with an additional specific 3'-exonuclease activity dependent on: i) the kind of pair: i.e. ds_r8oxoG:dC or ds_r8oxoG:dA; ii) salt concentrations (i.e. Mg^{2+}); iii) the presence of the 33 N-terminal domain.

Overall, our data demonstrate that BER enzymes but not RER are involved in the processing of non-canonical rNMPs, such as abasic or oxidized, incorporated in DNA. Because the repair pathway catalyzed by APE1 towards the oxidized rG is fundamentally different from that of BER, since no bases are excised, we can conclude that the observed effect could be ascribed to an alternative damage-specific endonuclease initiated repair pathway, previously referred to as either alternative excision repair (AER) or nucleotide incision repair (NIR) (as reviewed in [142, 143]).

3.4.6 APE1 and RNase H2 do not biochemically and functionally interact

To support the functional independence between BER and RER, we tested whether APE1 and RNase H2 proteins functionally interact. We tested whether this hypothesis was confirmed in HeLa cells transfected with siRNAs specific for APE1 and RNase H2 mRNAs to knock down the corresponding endogenous proteins. Whole cell extracts were prepared as explained in Materials and Methods section, and Western blotting analysis was used to check the effective protein down regulation (**Figure B.7a**). Then, we checked the ability of APE1 and RNase H2 from cell extracts to recognize the ds_r8oxoG:dC substrate. We performed EMSA with cell extracts from control (Scramble) and APE1-kd (siAPE1) or RNase H2-kd (siRNase H2) cells. As demonstrated in **Figure B.7b**, incubation of cell extract from control cells displayed a retarded band containing APE1-complex, as also demonstrated by supershift EMSA experiments with anti-APE1 specific antibody (**B.7c,d**). The intensity of the retarded complex was decreased upon APE1-kd (siAPE1) (**Figure *B.7b,*** lane 4) and upon the double APE1/RNase H2-kd (siAPE1 + siRNase H2) (*lane 6*) but not upon RNase H2 silencing alone (siRNase H2) (*lane 5*), confirming that APE1 is

involved in a protein complex able to recognize r8oxoG damage in which RNase H2 is not present.

3.5 Discussion

Increasing body of evidence supports the notion that incorporation of rNMPs in DNA is a frequent phenomenon, having profound detrimental effects on genome stability of both prokaryotes and eukaryotes [6, 38, 70, 144]. In humans, as well as in yeast and bacteria including Archaea, the main processing pathway responsible for repairing of these lesions is the RER pathway, which involves the RNase H2 enzyme [1, 145]. RNase H2 importance in higher organisms is testified by its essentiality for embryonic development in mouse [146]. Moreover, RNase H2 mutations in humans are causally linked to the onset of AGS, a rare autoimmune inflammatory disease [34, 147]. It can be hypothesized that among the many millions rNMPs that are introduced in the mammalian genome per cell cycle [27], not only canonical rNMPs are incorporated but also damaged rNMPs (such as abasic and oxidized). Indeed, like deoxyribonucleotides, rNMPs are also susceptible to oxidative insults [24, 148], and a significant generation of abasic sites formation has been demonstrated upon RNA oxidation and alkylation [149]. While the role of RNase H2-initiated RER mechanism of DNA repair in recognizing and cleaving rNMPs embedded in DNA is well established [27, 44], nothing is known regarding the DNA repair pathways involved in the removal of damaged rNMPs.

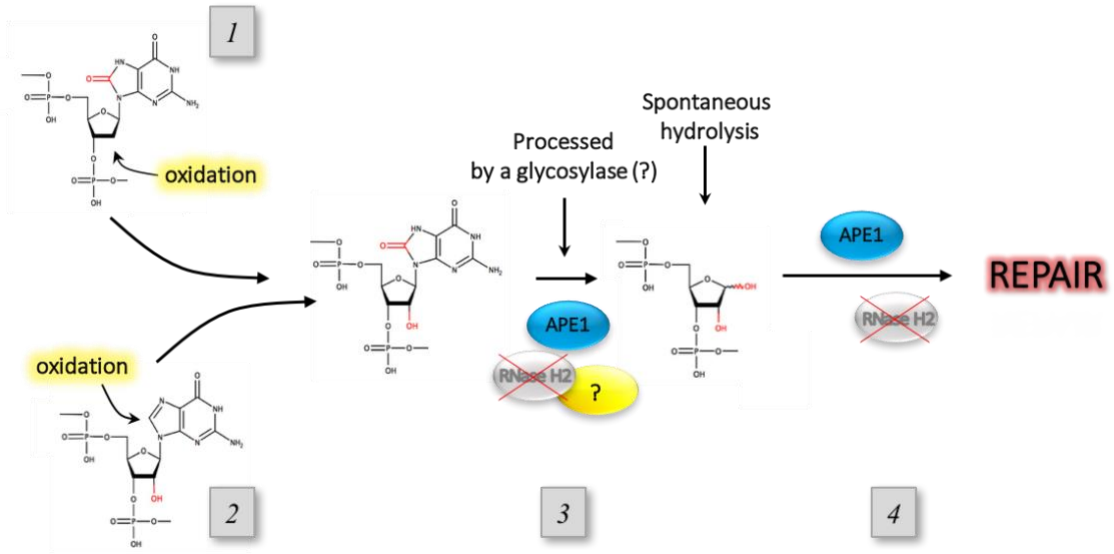
BER is the main mechanism coping with the repair of non-distorting single-base lesions, such as abasic sites and oxidized bases [64]. Interestingly, emerging literature, including ours, pointed to a new function of BER in RNA quality control surveillance and RNA-

decay with the functions of SMUG1, PARP1 and APE1 in RNA processing [66, 150]. At present, however, there is no evidence that BER may cope with the removal of rNMPs from DNA. Identifying whether BER may target unmodified and/or modified rNMPs in DNA is important to better understanding the mechanism of genotoxicity of oxidative stress and the impact of BER defects in human disease, cancer mechanisms, and for the development of new anticancer strategies.

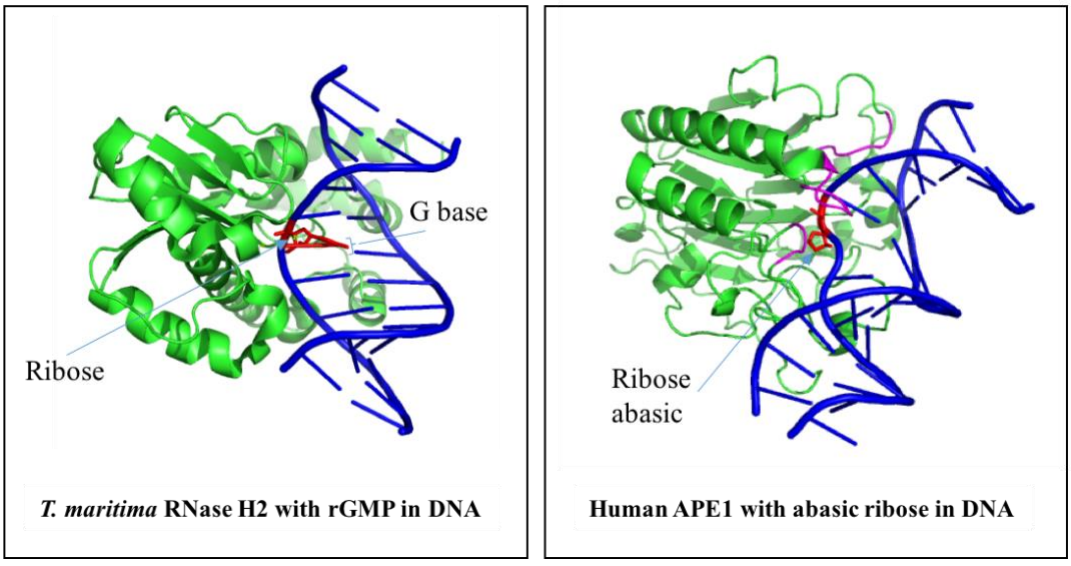
In this work, we demonstrate that an abasic rNMP embedded in DNA is targeted by APE1 of BER rather than RER in eukaryotic systems (**Figure 3.11a**). We have found that eukaryotic RNase H2 enzymes from yeast, mouse and human, and from human cell extracts are unable to process abasic rNMPs in DNA, whereas recombinant human APE1 is able to efficiently cleave this type of damage (**Figures 3.2** and **3.3**). APE1 processes the rAP site as efficiently as the canonical deoxy-abasic site as measured by kinetic data (**Table 3.1**). Data using catalytic inactive APE1 mutants (E96A) clearly demonstrate that the endonuclease active site of APE1 is required to perform the endoribonuclease activity on a ribose abasic site in dsDNA. Differently, the cleavage activity of the 33N-terminal truncated mutant, which does not impact the catalytic function of the enzyme but is involved in the release of the product upon cleavage [130], is comparable to that of wild-type APE1. This last result demonstrates that the unstructured N-domain is dispensable for the enzymatic reaction on the abasic ribonucleotide site. Moreover, using HeLa nuclear cell extracts that were siRNA-depleted of APE1 or RNAse H2 proteins, we showed that the processing activity of the abasic rNMP in DNA depends only on the presence of APE1 and not on RNase H2. These results highlight a new role of APE1 in repairing abasic

rNMPs embedded in DNA, demonstrating that the catalytic site of APE1 and the mechanism of product release is similar to that of the canonical deoxy-substrate.

a



b



T. maritima **RNase H2 with rGMP in DNA**

Human APE1 with abasic ribose in DNA

Figure 3.11 Model for repair of oxidized rNMPs and rAP sites embedded in DNA by APE1. (**a**) 8oxo-ribonucleotides could be generated in the cellular nucleotide pool or even when they are already incorporated in DNA as a result of an oxidation of the sugar (*1*) and/or the base (*2*). After the evidence that human RNase H2 is not able to process an oxidized rNMP embedded in DNA, we found that APE1 shows a weak but significant activity on it (*3*). Similarly, RNase H2 does not process an abasic rNMP embedded in DNA, which could be generated spontaneously or by the r8oxoG processing, and again APE1 possesses a strong activity on this type of damage (*4*). Because APE1 activity on 8oxo-ribonucleotides in DNA is low, we hypothesize that other proteins (some glycosylases?) may participate in their repair. (**b**) Structural models with the active site of RNase H2 and APE1 with an rNMP or an abasic rNMP. *T. maritima* (left) and human APE1 (right) is in a complex with DNA having a single rNMP or single abasic residue, respectively. DNA is indicated in blue, while the single rNMP and the abasic residues are shown in red as sticks. The arrows points towards ribose or G base or abasic sites in the panels. Proteins are shown in green except for three of the several regions on APE1 that engulf the abasic sugar which are in magenta. PDB for RNase H2 is 303-F; for APE1 as 1DEW.

Abasic rNMPs embedded in DNA may be generated by spontaneous hydrolysis or by the action of an unknown glycosylase on oxidized rNMPs, such as r8oxoG. Furthermore, a potentially significant, yet poorly characterized, source of rNMPs incorporated in DNA is the oxidative stress. rNMPs were shown to form during oxidative DNA damage both *in vitro* and *in vivo* [24]. Therefore, it is also possible that abasic and oxidized DNA is converted into RNA. It was estimated that spontaneous depurination occurs 1,000 times slower in RNA than DNA [151]. For example, the rate of depurination in DNA under physiological conditions is estimated to be 1 out of 100,000 purines every cell cycle. This rate gives 10,000 abasic sites per day in human cells [26]. Considering the remarkable abundance of rNMPs in DNA, which could be as many as 600,000 rNMPs in budding yeast genomic DNA, and therefore a factor of 250 higher in mammalian genomic DNA (150,000,000) [27], i.e. >300 millions in the human diploid DNA per cell cycle, it is not unrealistic to anticipate that cells may contain a non-negligible number of abasic rNMPs

in DNA, or oxidized rNMPs. Interestingly, abasic RNA results significantly more stable than abasic DNA, suggesting that specific enzymatic mechanisms should exist *in vivo* to cope with this harmful lesion [152]. In addition, in conditions of oxidative DNA damage, such as in cancer cells, the likelihood of such base modifications can increase. Moreover, recent studies already suggested that r8-oxoGTP is formed *in vivo* under oxidative stress conditions and may be incorporated during replication into DNA by *S. pombe*, *M. smegmatis* and human [118-120]. Moreover, the introduction of r8oxoG in DNA can be catalyzed by human DNA Pol β [120]. However, the level and function of Pol β in cells are highly regulated by complex signaling mechanism and interactome networks [153]. Therefore, the probability of r8oxoG incorporation into genomic DNA is still question under debate and theoretical estimation of the possible number of these incorporations needs experimental proves. Being spontaneous formation of rAP sites a rare event, it is a key question to find activities producing rAPs following processing of oxidized rNMPs in DNA. In addition, a possible RER and BER involvement in removal of r8oxoG from DNA has been proposed in other reports [120, 154]. Work in our laboratory is ongoing along these lines to address these fundamental issues. Considering these observations, we focused our attention on which BER protein, if any, may be involved in recognition and cleavage of oxidized rNMPs (r8oxoG) embedded in DNA. First, we tested the RNase H2 activity on r8oxoG substrate and our data clearly demonstrate that RER is not involved in processing of oxidized rNMPs embedded in DNA (**Figure 3.7**). Based on these findings, we explored a potential role of the BER pathway. First, we showed that the human OGG1, the main glycosylase enzyme able to recognize and repair oxidized dG through its lyase and glycosylase activities, has neither a lyase nor a glycosylase activity on an oxidized rG site

embedded in a DNA substrate, despite its ability to efficiently bind the oxidized substrate (**Figure 3.8**), in agreement with recent findings [154]. Interestingly, we discovered that APE1 has a weak endoribonuclease activity on r8oxoG site embedded in a DNA substrate, and shows a 3'-exonuclease activity (**Figures 3.9 and 3.10**), similarly to the 3'-exonuclease activity on DNA demonstrated previously [133, 135]. In line with previous results, the 3'-exonuclease activity of APE1 is strictly dependent on Mg^{2+} concentration and on the presence of the first 33 amino acids. The importance of the N-terminal domain is to be attributed to different reasons: i) it bears the majority of the positive charges of APE1; ii) is the target of the main post-translational modifications of the protein (i.e. acetylation, ubiquitination, proteolysis); iii) is involved in modulating the interaction with different protein partners, and finally iv) may modulate the catalytic rate, probably acting on the k_{off} of the catalytic reaction due to increased speed of product release [67, 130, 155]. These unexpected results, which suggest that APE1 3'-exonuclease activity strongly depends on the electrostatic interaction of APE1, involving its unstructured N-terminal domain, with the substrate [133], may be explained on the basis of the previously characterized ability of APE1 to process some particular structured RNA species in a specific manner [67].The activity of APE1 on r8oxoG embedded in DNA does not hide the possibility that a similar activity could be exerted on RNA molecules too. This finding could represent the first demonstration of an enzyme able to recognize and process oxidized RNA [155]. To date, RNA oxidation has been shown to exert detrimental physiological effects and to be a common feature in different human pathologies ranging from ageing to neurodegenerative and cancer diseases [156]. For instance, oxidized RNA [157] or RNA containing abasic sites [152] show inhibitory effects on reverse transcriptase activity,

whereas oxidized mRNA [158, 159] or mRNA with abasic sites [160] exhibit compromised translation activity as well as translation fidelity [161]. The weak endo- and 3'-exonuclease activities on r8oxoG- containing substrate and their dependence on Mg^{2+}-concentrations and on the presence of the first N-terminal domain residues of APE1 are fully in agreement with the previously described nucleotide incision repair (NIR) function by APE1 on several oxidized substrates, such as: 5,6-dihydro-2'-deoxyuridine, 5,6-dihydrothymidine, 5-hydroxy-2'-deoxyuridine, 5-hydroxycytosine [125-127]. Notably, the limited activity obtained in the experimental conditions we used (also after changing the pH conditions, data not shown) are in agreement with previous reports on the NIR function by APE1 on some particular substrates such as the αdG:dG and the 5OH-dC:dG [126]. In addition, the biochemical characterization through MALDI-MS and HPLC analyses we performed (**Figure B.5a,b**) may be suggestive for the existence of an equilibrium between different conformational species of r8oxoG dsDNA, excluding any possible bias due to contaminant present in the oligonucleotide used for the assays. Therefore, these findings underscore the importance of identifying the enzyme(s) responsible for the recognition and efficient processing of the r8oxoG substrate, in order to further extend our studies and our understanding of this hot scientific topic.

We observed that OGG1 is unable to process the r8oxoG substrate, while being perfectly able to specifically bind it, similar to what recently published by Sassa *et al*.,[154]. At present, it is possible to speculate that though the base-flipping occurs, the enzymes is unable to hydrolyze the N-glycosidic bond and has no lyase activity maybe as a consequence of the steric hindrance with the 2'OH of the ribose which renders the C1' unavailable for the nucleophilic attack by the catalytic site. Differently from our results,

Sassa *et al*., found no enzymatic activity by APE1. This discrepancy with our results may be due to the different experimental conditions for the enzymatic assays, i.e. higher Mg^{2+} concentrations and the use of a small amount of EDTA, both aspects already demonstrated to strongly affect the APE1 enzymatic activity on the r8oxoG substrate. A comparative experiment we performed (**Figure B.8**) was indeed supportive of this hypothesis, reinforcing the importance of the experimental conditions when studying the non-canonical functions of APE1 protein. Interestingly, Sassa *et al* showed that the commercially available prokaryotic RNase HII preserves the ability to remove an oxidized rNMP in a DNA duplex. Contrary to these results, our findings show that eukaryotic RNase H2 is completely inactive on a substrate containing an oxidized rNMP. These data suggest that the ability to process r8oxoG in DNA has been lost during evolution and deserves further studies.

To explain why APE1 recognized the abasic rNMP in DNA and RNase H2 did not, we compared the model structures of DNA with an rNMP or a rAP site in the RNase H2 and APE1 active site, respectively (**Figure 3.11b**). RNase H2 recognizes the RNA-DNA junction with the substrate participating in catalysis. Prior to incision by RNase H2, the rNMP base is hydrogen bonded to the complementary DNA strand base [45]. If the rNMP is abasic, there is no hydrogen bonding to stabilize the complex required for RNase H2. Rather, an orphan base on the complementary DNA strand is present. We hypothesized that the lack of the hydrogen bonding between the abasic rNMP and the opposite deoxyribonucleotide interferes with the capacity of RNase H2 to recognize an abasic rNMP and cleave it. Thus, the role of recognizing and cleaving abasic rNMPs is not specific of RNase H2. Differently from RNase H2, APE1 specifically recognizes and cleaves an

abasic distortion in DNA, and basically engulfs the sugar-phosphate further distorting the DNA [162], as seen in **Figure 3.11b**. Here, we predicted that the ribose extra OH would have only minor influence on the structure. Therefore, this could explain why the abasic rNMP, like an abasic deoxyribonucleotide in DNA, was efficiently cleaved by APE1.

Experiments are underway in order to address if other glycosylases may process oxidized rNMPs in order to generate abasic rNMPs, which are then efficiently processed by APE1. We previously demonstrated that APE1-defective cells have increased oxidized rRNA content upon oxidative stress [67]. This result has now a molecular explanation in the observed endoribonuclease activity of APE1 over r8oxoG containing oligonucleotides and will deserve further attention in our future studies. Regarding the formation of abasic ribonucleotides in DNA, this is still a matter of debate. The existence of specific N-ribohydrolases, including the toxin ricin, has been already documented [163] to be able to generate abasic rNMPs in RNA molecules, besides spontaneous generation [149]. A role for the YB-1 protein in recognizing oxidized ribonucleotides sites in RNA has also been hypothesized [164], but no specific enzymatic mechanisms able to remove the oxidized base has been described, yet. The accumulation of the r8oxoG substrates, which occurs on RNA upon silencing of APE1 expression, may thus be explained under the assumption that enzymatic removal of oxidized rNMPs may represent the limiting step in the process. Besides its direct activity on r8oxoG, APE1 could be stimulated by a glycosylase activity allowing a faster turnover as demonstrated for DNA substrates [165]. Work is in progress along these lines to better inspect this mechanism and the putative glycosylase enzymes involved. Moreover, since APE1 is overexpressed in different types of cancer, such as ovarian, gastro-esophageal, pancreaticobiliary, lung and breast cancers [166, 167], it would

be interesting to determine whether any correlation exists between its expression level and presence of modified rNMPs in cancers.

3.6 Acknowledgements

We thank the Tell and Storici laboratories for constructive feedback. They are also grateful to the funding sources listed below for supporting this work. The authors thank Bret Freudenthal for his fruitful comments during manuscript preparation. This work was supported in part by the National Institute of Health, NIH [R01ES026243-01 to F.S.], the Howard Hughes Medical Institute Faculty Scholar [55108574 to F.S.], the Italian Association for Cancer Research, AIRC [IG2013-14038 to G.T.]; and the Intramural Research Program of National Institute of Child Health and Human Development [to R.J.C.].

CHAPTER 4. CAPTURE OF RIBONUCLEOTIDES IN YEAST GENOMIC DNA USING RIBOSE-SEQ

The procedure described in Chapter 4 consists of the publication accepted in *Methods in Molecular Biology 'Yeast Systems Biology. Methods and Protocols' (2nd ed),* Springer, edited by Steve G. Oliver and Juan Castrillo.

Balachander, S.[1†], Yang, T.[1†], Newnam, G.[1], El-Sayed, W.M.M.[1], Koh, K.D.[2], & Storici, F.[1]

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA.
[2]Department of Medicine, University of California, San Francisco, California, USA.
[†]Equal Contribution

4.1 Abstract

Experiments conducted in yeast cells have recently shown abundant presence of ribonucleotides (rNMPs) embedded both in nuclear and mitochondrial DNA. Indeed, rNMPs are the most frequent, non-standard nucleotides found in cellular DNA. rNMPs have a highly reactive 2'-hydroxyl group in the ribose sugar that gives rise to genome instability by altering the structure, function, and properties of DNA. In order to profile rNMPs embedded in yeast genomic DNA, as well as any other genomic DNA of interest, we developed 'ribose-seq'. Ribose-seq utilizes *Arabidopsis thaliana* tRNA ligase (AtRNL), which enables ligation of 2'-phosphate termini of DNA molecules terminating with an rNMP to the 5'-phosphate end of the same DNA molecules. Thus, a unique feature of ribose-seq is its capacity to specifically and directly capture the rNMPs present in DNA. Here we describe how ribose-seq is applied to yeast *Saccharomyces cerevisiae* DNA to capture rNMPs that are incorporated in the yeast genome and build libraries of rNMP incorporation for high-throughput sequencing. We also provide advancements over our original ribose-seq protocol, which are described at the end of the Introduction section while the specific details are provided in the methods part of this chapter.

4.2 Introduction

Recent mapping of ribonucleotides (rNMPs) embedded in yeast genomic DNA to single-nucleotide resolution has unraveled initial important aspects of rNMP incorporation in DNA, suggesting that rNMPs could have a profound impact on DNA metabolic functions [168]. Our group developed ribose-seq technique to map rNMPs in DNA and applied it to

determine the profile of rNMP incorporation in yeast nuclear and mitochondrial DNA [70]. We and others have found that rNMPs are widely distributed in the yeast genome both in nuclear and mitochondrial DNA; however, sites of rNMPs are not random [68-71]. Using yeast cells deficient for ribonuclease H2, which is the major enzyme responsible for the removal of rNMPs in DNA, a significant bias of rNMP incorporation was found between leading and lagging strand of DNA replication sites in yeast DNA. A predominant presence of rNMPs in the leading strands was revealed in all reports that analyzed rNMPs in budding or fission yeast DNA [68, 71, 169]. Moreover, rNMPs were found to have hotspots of incorporation in yeast nuclear and mitochondrial DNA, and abundant incorporation was observed at G/C rich regions in mitochondrial DNA of *S. cerevisiae* cells [70].

rNMPs have been found to be the most abundant non-standard nucleotides in DNA of any cell type that has been analyzed so far [6]. A major source of rNMP incorporation in the genome is by DNA polymerases. Many DNA polymerases have been shown to incorporate rNMPs in DNA, including yeast replicative polymerase Pol α, δ, and ε [27]. The presence of rNMPs in DNA can have positive and negative consequences. The presence of two rNMPs at the mating type locus is believed to activate mating type switching in fission yeast [42]. In addition, due to the prevalence of rNMPs on the leading versus the lagging strand of DNA replication, rNMPs can act as a strand discrimination signal for eukaryotic mismatch repair [40, 170]. On the other hand, due to the presence of the highly reactive 2'-hydroxyl group in the ribose sugar, rNMPs in DNA can cause genomic instability of the cells [2]. rNMPs have also been shown to modify the structural, and mechanical properties of DNA [37]. The presence of rNMPs results in increased mutation rates and chromosomal

abnormality [53, 171]. Moreover, accumulation of rNMPs was found to be associated with defects in RNase H2 in cells of Aicardi Goutieres syndrome patients [55].

In order to better understand the significance and the consequences of rNMP incorporation in DNA, it is important to uncover the genomic sites of rNMPs incorporation in different genetic backgrounds and in cells grown in different conditions. The ribose-seq protocol is based on the activity of a special ligase, the tRNA ligase of *Arabidopsis thaliana* (AtRNL), which we showed to ligate 5'-phosphate end of DNA to a 2'-phosphate end of a DNA terminated with an rNMP [70]. Alkali denatures DNA strands, hydrolyzes RNA into ribonucleotides with a 2',3'-cyclic phosphate, 3' phosphate or 2'-phosphate end, and cleaves at an rNMP in DNA generating a ssDNA fragment terminated by a 2',3'-cyclic phosphate, 3' phosphate or 2'-phosphate. Upon alkali treatment, the AtRNL converts 2',3'-cyclic phosphate ends into a 2'-phosphate ends and attaches all terminal rNMP sites with a 2'-phosphate end of a DNA fragment with unknown sequence to a 5'-phospate end of a DNA adaptor for Illumina sequencing. By exploiting AtRNL activity, the ribose-seq protocol cannot capture unligated RNA primers of Okazaki fragments or any site of single or double-strand break in DNA, because in none of these cases the DNA displays a 2'-phospate end upon alkali treatment [70]. Thanks to this unique feature, the ribose-seq technique can be applied to study and characterize the pattern of rNMP incorporation in practically any cell type of interest, including yeast cells of any species, ploidy, and genetic background. Moreover, ribose-seq can also be applied to profile rNMPs in yeast or other cell types that are grown in any desired conditions, that have been treated using any mutagen, drug, DNA damaging agent, toxicant of interest, or exposed to any chosen environmental factor [70].

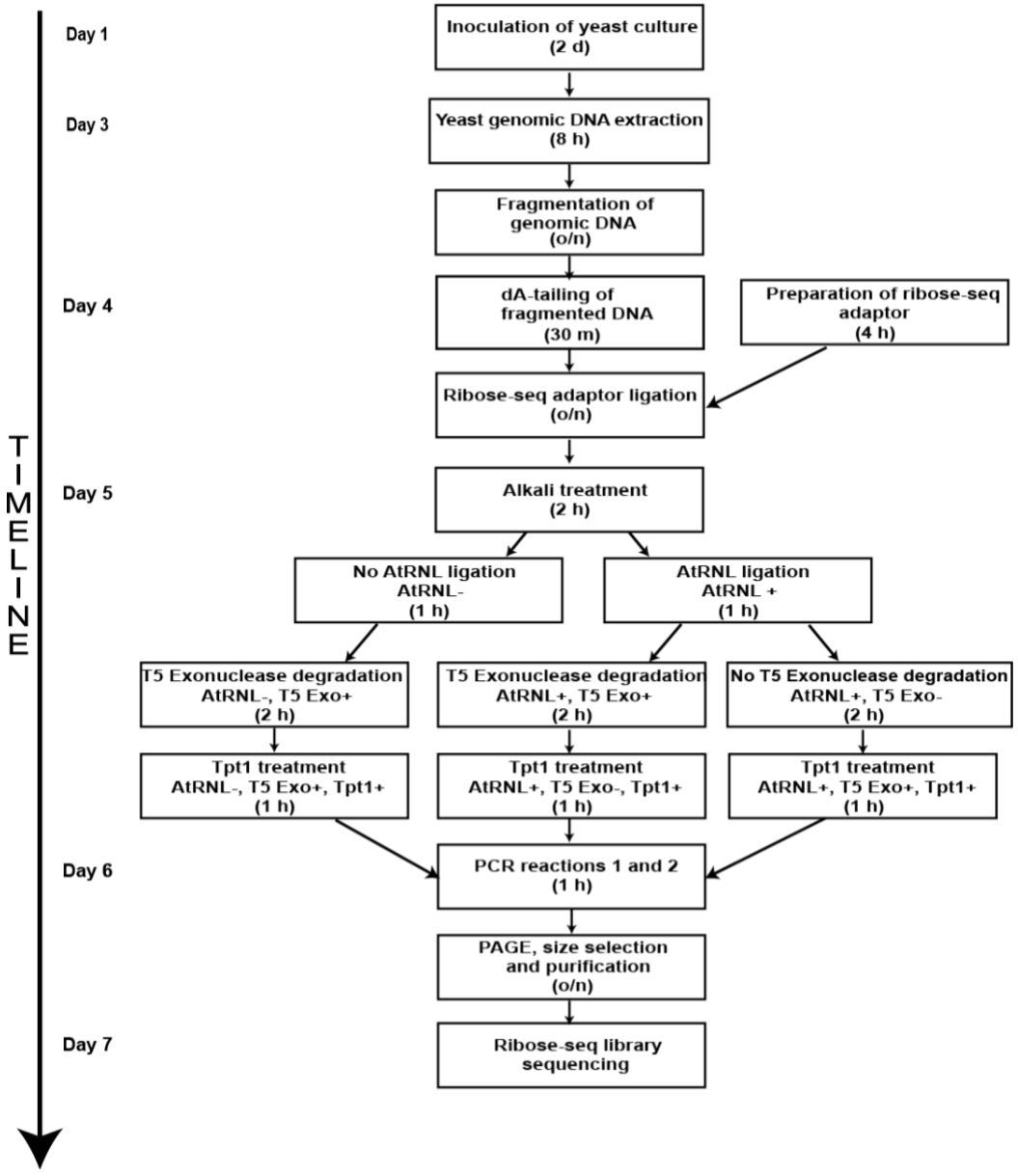


Figure 4.1 Timeline of ribose-seq protocol. A yeast culture is grown in YPD liquid medium at 30 °C for 2 days. On day 3, genomic DNA is extracted from the yeast cells and fragmented by blunt-ended restriction enzymes. On day 4, ribose-seq adaptors are prepared, and the fragmented double-strand DNA (dsDNA) undergoes dA-tailing reaction. This is followed by ribose-seq adaptor ligation on day 4. On day 5, the dsDNA fragments ligated to the adaptors are treated with alkali. The alkali treated DNA is split in two samples. One sample is incubated in the presence of AtRNL (AtRNL+), the other not (AtRNL-). The AtRNL+ sample is then treated with T5 Exonuclease (T5 Exo+) or without (T5 Exo-) to degrade the unligated single-strand DNA (ssDNA). The AtRNL- sample is treated only with T5 exonuclease. All samples are further treated with Tpt1 to remove the 2' phosphate. On day 6, the Tpt1-treated samples are amplified by PCR in two rounds (PCR round 1 using PCR.1 and PCR.2 primers and PCR round 2 to attach the i5 and i7 indexes

for Illumina sequencing). After that, the PCR products are loaded onto a non-denaturing polyacrylamide gel (PAGE) for visualization of the ribose-seq library and gel purification before sequencing. From the beginning to the end, ribose-seq library preparation can be performed.

Figure 4.1 provides a scheme of the steps and the timeline of the procedure to build ribose-seq libraries from yeast genomic DNA. Briefly, yeast cells of the chosen strain are incubated in the desired conditions and are collected to reach approximately 10^{10} cells (**Figure 4.2**). The yeast cells are lysed to extract at least ~40 μg of yeast genomic DNA, or more. The DNA is then fragmented using restriction enzymes to produce blunt-ended fragments with an average size of ~450 base pairs (bp) in length (**Figure 4.2**). This is followed by dA-tailing and ligation to a double-strand adaptor that contains a molecular barcode (**Figure 4.2**). The fragments are treated using alkali to denature the DNA strands and cleave rNMP sites by exposing the 2',3'-cyclic phosphate or 2'-phosphate at the rNMP termini of the DNA fragments, which are recognized by AtRNL (**Figure 4.3**). By utilizing AtRNL, we ligate the 2'-phosphate ends of each rNMP-terminated single-stranded (ss) DNA fragment to its opposite 5'-phosphate end, thereby resulting in single-stranded circular DNA forms that contain the embedded rNMP next to the adaptor for Illumina sequencing. This is followed by degradation of the unligated ssDNA fragments using T5 Exonuclease, and treatment with 2'-phosphotransferase Tpt1 to remove the 2'-phosphate at the ligation junction in the circular forms containing the rNMP (**Figure 4.3**). The circular molecules are then PCR-amplified to generate a ribose-seq library that contains the rNMPs and their upstream sequences (**Figure 4.4**). The PCR product is then loaded into a PAGE gel to detect the libraries together with control samples (**Figure 4.5**). A smear

corresponding to the ribose-seq library after 15 cycles of PCR 1 and 13 cycles of PCR 2 is seen in lane 5 of the gel image shown in **Figure 4.5**. No smear is seen after the same number of PCR cycles for the control AtRNL- sample, lane 3 in the same gel image. To reduce the number of PCR 1 and 2 cycles, we run 10 cycles for PCR1 and 7 cycles for PCR 2 for the extraction of the ribose-seq library. Because this lower number of cycles often does not allow detection of the ribose-seq library product on PAGE (lane 5 in **Figure 4.6**), we also perform PCR 2 with 15 cycles to visualize the ribose-seq library product if present (lane 6 in **Figure 4.6**). The ribose-seq product of several reactions generated using PCR 1 with 7 cycles and PCR 2 with 10 cycles is then loaded in multiple wells of another PAGE gel to extract the ribose-seq library within a size-range of 200-700 bp to exclude any primer dimers and long products that are not proficient sequencing (**Figure 4.7**). Overall, the ribose-seq protocol for yeast genomic DNA takes 7 days. The amplified DNA library is sequenced and subsequently analyzed using our Ribose-Map bioinformatics toolkit [172].

In this protocol, we provide a detailed procedure of ribose-seq starting from extraction of yeast genomic DNA (gDNA) to the preparation of libraries for sequencing. We also include modifications to the original-described ribose-seq method [70]. (i) We redesigned the molecular barcode-containing adaptor by making it shorter and removing overlapping sequences. (ii) We cut the genome of interest in smaller fragments (~450 bp). (iii) We introduced two rounds of PCR and reduced the PCR cycle number. (iv) We cut and purify a specific size-range of the ribose-seq library from the non-denaturing gel to eliminate any primer dimers generated during PCR and long products that are not proficient for sequencing because these have low affinity to the flow cells in the Illumina sequencing

systems. By including these improvements, the efficiency of ribose-seq increased up to a factor of 1,000.

4.3 Materials

Prepare all solutions using RNase/DNase free water. Store all solutions and reagents at room temperature (unless indicated otherwise).

4.3.1 Yeast genomic DNA extraction

1. YPD liquid (for 1 L)–10 g of yeast extract, 20 g of bacto peptone and 20 g of dextrose. Autoclave at 121 ℃ for 20 min (*See* **Note 1)**.
2. 250 mL glass flasks.
3. 250 mL Nalgene plastic flasks.
4. Buffer A–Dissolve 182.2 g of sorbitol in 600 mL of water. Add 200 mL of a 0.5 M Na_2EDTA (pH 8.0) solution. Add 1 mL of B-mercaptoethanol (14.3 M). Finally adjust the volume to 1 L with water.
5. Zymolyase 20T (US Biological): Add 0.05 g from 20 KU stock.
6. RNase A.
7. Buffer B–Dissolve 76.42 g of guanidine HCl, 11.17 g of $Na_2EDTA{\cdot}2H_2O$, and 3.63 g of Tris base in 600 mL of water. Add 250 mL of 20% Tween-20 solution and 50 mL of 10% Triton X-100 solution. Use NaOH to adjust the pH to 8.0 and adjust the volume to 1 L with water.
8. Proteinase K powder (US Biological): Prepare a stock solution of 20 mg/mL.

9. Buffer C–Dissolve 43.83 g of NaCl and 10.46 g of MOPS in 800 mL of water. Use NaOH to adjust the pH to 7.0 and add 150 mL of isopropanol and 15 mL of 10% Triton X-100 solution. Adjust the volume to 1 L with water.
10. Buffer D–Dissolve 58.44 g of NaCl and 10.46 g of MOPS in 800 mL of water. Use NaOH to adjust the pH to 7.0 and add 150 mL of isopropanol. Adjust the volume to 1 L with water.
11. Buffer E–Dissolve 73.05 g of NaCl and 6.06 g of Tris base in 800 mL of water. Use HCl to adjust the pH to 8.5 and add 150 mL of isopropanol. Adjust the volume to 1 L with water.
12. 50 mL conical tubes.
13. QIAGEN Genomic-tip 500/G.
14. 1.5 mL Eppendorf tubes.
15. Isopropanol.
16. Cold 70% ethanol.
17. DNase/RNase-free water.
18. Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific).

4.3.2 Annealing of ribose-seq adaptor oligonucleotide

1. Adaptor.L oligonucleotide (HPLC-purified, 5′-P-NNC CGN NNN NNA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTT GAT GAT AGA TCC GTG TCG CAA CT -3′) (Integrated DNA Technologies) (*See* **Note 2**).
2. Adaptor.S oligonucleotide (HPLC-purified, 5′-P-GTT GCG ACA CGG ATC TAT CAA CAC T -Am-3′) (Integrated DNA Technologies) (*See* **Note 2**).
3. Annealing Buffer–500 mM Tris-HCl, pH 7.5, 2.5 M NaCl, 50 mM EDTA).

4. illustra MicroSpin G-25 Column (GE Healthcare Life Sciences).

4.3.3 Ribose-seq

4.3.3.1 Fragmentation of DNA

1. 40 μg of genomic DNA.
2. 10X NEBuffer 2 (New England Biolabs).
3. 20 U/μL DraI (New England Biolabs).
4. 20 U/μL EcoRV (New England Biolabs).
5. 5 U/μL SspI (New England Biolabs).
6. 10 U/μL AluI (New England Biolabs).
7. PCR Purification Kit (QIAGEN).

4.3.3.2 dA-Tailing and adaptor ligation

1. 10 mM dATP (Sigma-Aldrich).
2. 10X NEBuffer 2 (New England Biolabs).
3. 5 U/μL Klenow Fragment (3′→5′ exo–) (New England Biolabs).
4. PCR Purification Kit (QIAGEN).
5. 10X T4 DNA Ligase Reaction Buffer (New England Biolabs)
6. 400 U/μL T4 DNA Ligase (New England Biolabs).
7. Agencourt RNAClean XP (Beckman Coulter).

4.3.3.3 Alkali Treatment

1. 2 M NaOH (*See* **Note 3**).
2. 2 M HCl.

3. pH Litmus Paper.
4. Agencourt RNAClean XP (Beckman Coulter).

4.3.3.4 AtRNL-Ligation

1. 10 μM purified AtRNL protein (prepared in our laboratory).
2. 10X AtRNL Reaction Buffer–500 mM Tris-HCl (pH 7.5), 400 mM NaCl, 50 mM $MgCl_2$, 10 mM DTT, and 300 μM ATP.
3. Agencourt RNAClean XP (Beckman Coulter).

4.3.3.5 T5-Exnuclease Degradation

1. 10X NEBuffer 4 (New England Biolabs).
2. 10 U/μL T5 Exonuclease (New England Biolabs).
3. Agencourt RNAClean XP (Beckman Coulter).

4.3.3.6 Tpt1 Reaction

1. 20 μM purified Tpt1 protein (prepared in our laboratory).
2. 10X Tpt1 Reaction Buffer–200 mM Tris-HCl (pH 7.5), 50 mM $MgCl_2$, 1 mM DTT, and 4% Triton X-100.
3. 50 mM NAD^+ (Sigma-Aldrich)
4. Agencourt RNAClean XP (Beckman Coulter).

4.3.3.7 PCR Amplification of DNA libraries

1. PCR.1 oligonucleotide (Desalted, 5'-GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TTG ATA GAT CCG TGT CGC AAC-3') (Integrated DNA Technologies).

2. PCR.2 oligonucleotide (Desalted, 5'-ACA CTC TTT CCC TAC ACG AC-3') (Integrated DNA Technologies).
3. PCR.701.Index oligonucleotide (Desalted, 5'-CAA GCA GAA GAC GGC ATA CGA GAT **CGA GTA AT**G TGA CTG GAG TTC AGA CGT GT-3') (Integrated DNA Technologies) (*See* **Note 4**).
4. PCR.501.Index oligonucleotide (Desalted, 5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC A**CT ATA GCC** TAC ACT CTT TCC CTA CAC GAC-3') (Integrated DNA Technologies) (*See* **Note 5**).
5. 5X Q5 Reaction Buffer (New England Biolabs).
6. 2 U/μL Q5 High-Fidelity DNA Polymerase (New England Biolabs).
7. 10 mM dNTPs (New England Biolabs).

4.3.3.8 PAGE Visualization

1. Non-denaturing 6% Polyacrylamide (29:1) Mini-Gel
2. 10X TBE Buffer.
3. 6X Sucrose Loading Buffer – Add 0.4 g of Sucrose, 0.00025 g of Bromophenol blue, 0.00025 g of Xylene cyanol, and 120 μL of 500 mM EDTA (pH 8.0) and adjust the volume to 1 mL with water.
4. Quick-Load 50 bp DNA Ladder (New England Biolabs).
5. SYBR Gold Nucleic Acid Stain (Thermo Fischer Scientific).
    a. M NaCl.
6. 3-mL Syringe.
7. 15 mL conical tubes.
8. 0.2-micron filter.

9. Razor blade.

4.3.4 Equipment

1. New Brunswick Incubator Shaker I.
2. Eppendorf Centrifuge 5424.
3. Eppendorf Centrifuge 5810.
4. Beckman Coulter Avanti J-E Centrifuge.
5. Water bath.
6. Qubit 3.0 (Thermo Fischer Scientific).
7. NanoDrop Spectrophotometer (Thermo Scientific).
8. Applied Biosystems Thermal Cycler 2720.
9. Mini-PROTEAN Tetra Cell (BIO-RAD).
10. Gel Imaging System (UV-light).
11. 2100 Bioanalyzer Instrument (Agilent).
12. Illumina Sequencing machine.

4.4 Methods

All procedures are carried out at room temperature unless specified otherwise.

4.4.1 Yeast genomic DNA extraction

1. Inoculate 150 mL of YPD liquid medium with yeast strain into a 500 mL glass flask and shake vigorously at 30 °C for 2 days.

2. Transfer the culture to a 250 mL Nalgene polypropylene bottle and pellet the yeast cells by centrifuging at 5,000 x *g* at 4 °C for 5 min. Discard the supernatant.
3. Resuspend the pellet in 12 mL of DNase/RNase free water by vortexing and proceed to pellet the cells by centrifuging at 5,000 x *g* at 4 °C for 10 min. Discard the supernatant.
4. Resuspend the pellet in 12 mL of Buffer A by vortexing at maximum speed to homogenize the cells as thoroughly as possible.
5. Transfer the resuspended cells into a 50 mL conical tube and add 0.05 g of zymolase and incubate at 30 °C for 1 h (*see* **Note 6**).
6. Pellet the spheroblasts by centrifuging at 5,000 x *g* at 4 °C for 10 min.
7. Resuspend the pellet in 15 mL of Buffer B with 30 µL of RNase A and mix the tube thoroughly.
8. Add 400 µL of Proteinase K stock solution and incubate at 50 °C for at least 1 h.
9. Pellet the cell suspension by centrifuging at 5,000 x *g* at 4 °C for 15 min. Transfer the supernatant to a new 50ml conical tube.
10. Vortex the supernatant for 10 s and apply it to an equilibrated QIAGEN Genomic-tip 500/G. Allow the genomic tip to empty by gravity flow (*see* **Note 7**).
11. Wash the tip twice with 15 mL of Buffer QC.
12. Elute the genomic DNA by adding 15 mL of pre-warmed Buffer QF (*see* **Note 8**).

13. Briefly vortex the eluted genomic DNA and distribute 750 μL into 1.5 mL Eppendorf tubes. Add 525 μL of isopropanol to each tube.
14. Vortex the tubes briefly and incubate them at -80 °C for 5 min. After that, pellet the mixture at 15,000 rpm at 4 °C for 20 min (*see* **Note 9**).
15. Add 200 μL of cold 70 % ethanol into the tubes. Vortex the tubes briefly and centrifuge at 15,000 rpm at 4 °C for 10 min.
16. Remove the supernatant carefully by pipetting without disturbing the pellet and air-dry the pellets for 10 min.
17. Resuspend the DNA pellet with 15 μL of DNase/RNase free water and incubate the tubes at room temperature overnight.
18. Collect the eluted DNA into a 1.5 mL tube and measure the concentration of the DNA using Qubit 3.0 High-Sensitivity DNA kit (*see* **Note 10**).

4.4.2 Preparation of ribose-seq adaptor

1. Resuspend each oligonucleotide (oligo) of ribose-seq adaptor in DNase/RNase-free water to a concentration of 50 μM and 500 μM, respectively.
2. Set up the mixture for preparing ribose-seq adaptor as follows:

Annealing Buffer	3 μL
50 μM Adaptor.L	25 μL
500 μM Adaptor.S	12.5 μL
H_2O	9.5 μL
Total	**50 μL**

3. Perform annealing by heating the mixture to 95–100 °C and gradually cooling to room temperature until the temperature drops to 30–32 °C. The annealed ribose-seq adaptor is at a concentration of 25 μM (*see* **Note 11**).
4. Desalt the mixture by using a spin column. In our laboratory, illustra MicroSpin G-25 Column is utilized.
5. Use NanoDrop to quantify the amount of desalted double-stranded ribose-seq adaptor. Typically, the concentration of double-strand adaptors ranges from 10-13 μM. A concentration of 10 μM will be assumed for subsequent steps of the protocol.

4.4.3 Ribose-seq

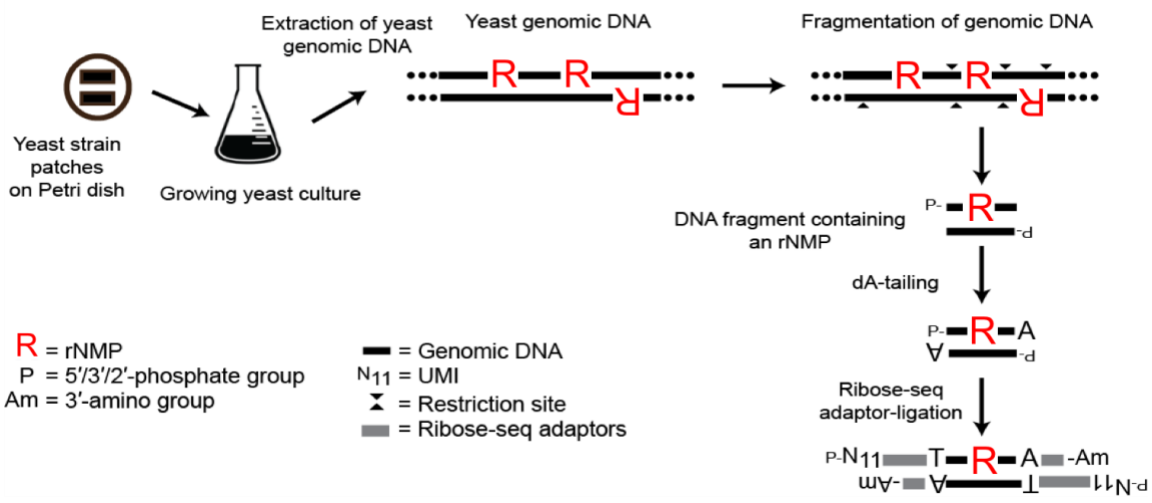


Figure 4.2 Scheme of the first steps of the ribose-seq protocol from the yeast culture to the adaptor ligation. Yeast cells are grown in 150 mL of YPD liquid medium for 2 days. Genomic DNA is extracted on day 3. Fragmentation of genomic DNA is conducted overnight at 37 °C. On day 4, the fragmented DNA is dA-tailed at 37 °C for 30 min, followed by ribose-seq adaptor ligation. The ligation reaction is performed at 15 °C overnight.

4.4.3.1 Fragmentation of genomic DNA

1. Set up 8 identical reactions of restriction enzyme digestion of genomic DNA as follows. In this protocol, we use AluI, DraI, EcoRV, and SspI (*see* **Note 12**).

10X NEBuffer 2	12 μL
Genomic DNA	5 μg (x μL)
10 U/μL AluI	3 μL
20 U/μL DraI	3 μL
20 U/μL EcoRV	3 μL
5 U/μL SspI	3 μL
H_2O	96-x μL
Total	**120 μL**

2. Incubate at 37 ℃ overnight.
3. Purify the fragmented DNA using a spin column. In this protocol, we use the QIAGEN spin column from their PCR Purification Kit. Two digestion reactions can be purified using a single column with elution volume of 30 μL per spin column.
4. Use Qubit 3.0 (dsDNA HS Assay Kit) to quantify the amount of fragmented DNA. Typically, the concentration of the resulting DNA is ~200 ng/μL, following the reaction conditions listed above. Approximately, 20 μg of fragmented genomic DNA is subject for subsequent steps of the protocol.

4.4.3.2 dA-Tailing and adaptor ligation

1. Set up 4 dA-tailing reactions as follows:

10X NEBuffer 2	5 μL

10 mM dATP	1 μL
200 ng/μL Fragmented DNA	5μg (x μL)
5 U/μL Klenow Fragment (3′→5′ exo–)	3 μL
H_2O	41-x μL
Total	***50 μL***

2. Incubate at 37 °C for 30 min and purify using QIAGEN spin column. Elute dA-tailed DNA with 30 μL of water (*see* **Note 13**).
3. Set up 4 sequencing adaptor-ligation reaction as follows:

10X T4 DNA Ligase Buffer	5 μL
10 μM ds-ribose-seq adaptor	5 μL
dA-tailed DNA	30 μL
400 U/μL T4 DNA Ligase	5 μL
H_2O	5 μL
Total	**50 μL**

4. Incubate at 15 °C overnight.
5. Purify using Agencourt RNAClean XP with elution volume of 30 μL.

4.4.3.3 Alkali-treatment

1. Set up 4 alkali-treatment reactions as follows:

2 M NaOH	7.5 μL
Adaptor-ligated	30 μL
H_2O	12.5 μL
Total	**50 μL**

2. Incubate at 55 °C for 2 h.

3. Neutralize with 2 M HCl to pH 7. Use pH Litmus Paper to check the pH. Typically, 7.5–8 µL is needed for neutralization.
4. Purify using Agencourt RNAClean XP with elution volume of 45 µL.
5. Heat the Alkali-treated DNA at 95 °C for 3 min to ensure complete denaturation of dsDNA and immediately chill on ice for 2 min.

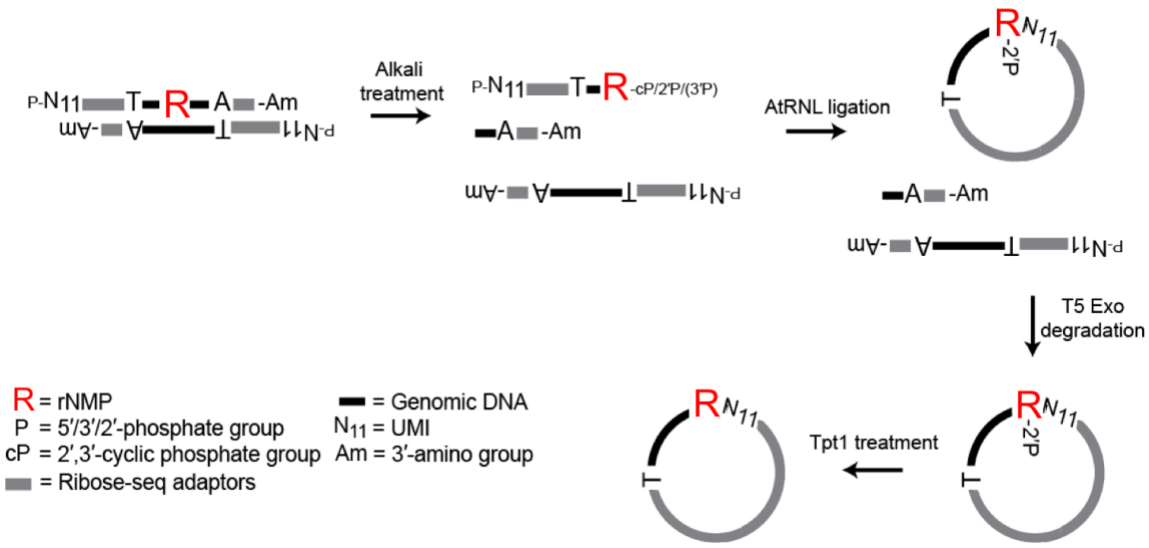


Figure 4.3 Scheme of the core steps of ribose-seq to capture the rNMPs embedded in DNA by alkali and AtRNL ligation. Adaptor ligated dsDNA is treated with 2M NaOH at 55 °C for 2 h to denature the dsDNA and cleave 3' to the rNMP sites. Alkali treatment generates 2',3'-cyclic phosphate, 3' phosphate or 2'-phosphate ends at each rNMP site. AtRNL does not recognize 3'-phosphate ends, but converts 2',3'-cyclic phosphate ends to 2'-phospate ends and ligates each 2'-phosphate end to its 5'-phospate end in an intramolecular-ligation reaction at 30 °C for 1 h. T5 exonuclease is then used to degrade the unligated ss linear fragments at 37 °C for 2 h. The resulting circular ssDNA fragments, each containing an rNMP, are treated with Tpt1 to remove the 2'-phosphate. Tpt1 reaction is carried out at 30 °C for 1 h. All of these reactions are performed in a single day (see **Figure 4.1**).

4.4.3.4 Self-ligation (circularization) of 2'-phosphate terminal of rNMP to 5'-phosphate of the DNA by AtRNL

1. Set up 4 reactions, one without AtRNL (AtRNL–) and 3 with (AtRNL+), as follows:

[AtRNL–]

10X AtRNL Reaction Buffer	2 μL
Alkali-treated DNA	6.3 μL
H_2O	11.7 μL
Total	**20 μL**

[AtRNL+]

10X AtRNL Reaction Buffer	2 μL
Alkali-treated DNA	12.6 μL
10 μM AtRNL	2 μL
H_2O	3.4 μL
Total	**20 μL**

2. Incubate at 30 ℃ for 1 h.
3. Purify using Agencourt RNAClean XP and elute with 30 μL for AtRNL- reaction and with 20 μL each for AtRNL+.

4.4.3.5 Removal of linear ssDNA

1. Set up 4 reactions, one without T5 Exonuclease (AtRNL+ T5Exo–) and one with T5 Exonuclease (AtRNL– T5Exo+) and 2 with T5 Exonuclease and AtRNL+ (AtRNL+ T5Exo–), as follows:

[AtRNL– T5Exo+]

10X NEBuffer 4	5 μL
AtRNL– DNA	30 μL
10U/μL T5 Exonuclease	5 μL
H_2O	10 μL
Total	**50 μL**

[AtRNL+ T5Exo–]

10X NEBuffer 4	5 μL
AtRNL+ DNA	20 μL
H_2O	25 μL
Total	**50 μL**

[AtRNL+ T5Exo+]

10X NEBuffer 4	5 μL
AtRNL+ DNA	20 μL
10U/μL T5 Exonuclease	5 μL
H_2O	20 μL
Total	**50 μL**

2. In order to maximize the amount of DNA that is available for a ribose-seq library, AtRNL– T5Exo– sample is omitted.
3. Incubate at 37 ℃ for 2 h.
4. Purify using Agencourt RNAClean XP with elution volume of 26 μL

4.4.3.6 Removal of 2'-phosphate

1. Set up 4 reactions with Tpt1, one for AtRNL– T5Exo+, one for AtRNL+ T5Exo–, and 4 for AtRNL+ T5Exo+ products, as follows:

[Tpt1+]

10X Tpt1 Reaction Buffer	4 μL
50 mM NAD^+	8 μL
DNA	26 μL
20 μM Tpt1	2 μL
H_2O	10 μL
Total	**40 μL**

2. Final reaction concentration of Tpt1 is 1 μM. DNA indicates either AtRNL– T5Exo+, AtRNL+ T5Exo–, or AtRNL+ T5Exo+ product.
3. Incubate at 30 °C for 1 h.
4. Purify using Agencourt RNAClean XP with elution volume of 30 μL.

4.4.4 PCR

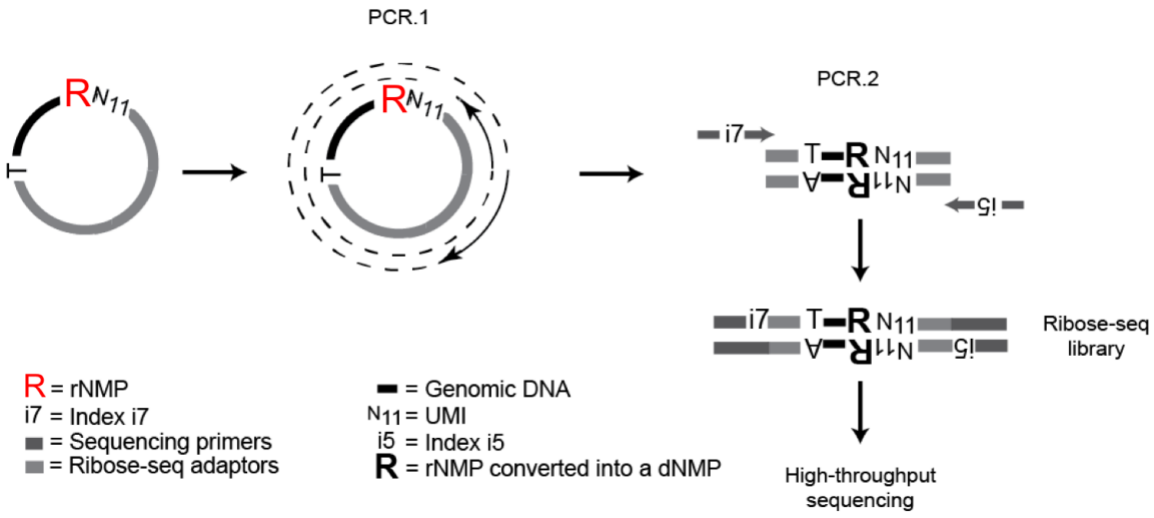


Figure 4.4 PCR reactions of ribose-seq. PCR amplification of circular ss products each containing an rNMP is performed by using two rounds of PCR. PCR round 1 is performed to amplify the ribose-library and to introduce the sequences of Illumina TruSeq index primers, and PCR round 2 is performed to attach specific TruSeq indexes i7 and i5.

4.4.4.1 PCR 1

1. Set up first PCR round 1 to amplify and introduce the sequences of Illumina TruSeq Index primers. To run 4 PCR reactions, one without template DNA (Primers only) and three with template DNA (AtRNL– T5Exo+ Tpt1+, AtRNL+ T5Exo– Tpt1+, or AtRNL+ T5Exo+ Tpt1+ products, as follows:

   [Primers-only]

5X Q5 Reaction Buffer	10 µL
10 mM dNTPs	1 µL
10 µM PCR.1	2.5 µL
10 µM PCR.2	2.5 µL

Q5-HF DNA Polymerase	0.5 μL
H_2O	33.5 μL
Total	**50 μL**

[Template+]

5X Q5 Reaction Buffer	10 μL
10 mM dNTPs	1 μL
10 μM PCR.1	2.5 μL
10 μM PCR.2	2.5 μL
Template DNA	x μL
Q5-HF DNA Polymerase	0.5 μL
H_2O	33.5-x μL
Total	**50 μL**

2. Template DNA indicates either AtRNL– T5Exo+ Tpt1+, AtRNL+ T5Exo– Tpt1+, or AtRNL+ T5Exo+ Tpt1+ products and use 15 μL, 5 μL, and 30 μL respectively.
3. Run PCR with the following settings:

Cycle Step	Temperature	Time	Cycles
Initial Denaturation	98 °C	30 s	1
Denaturation Annealing Extension	98 °C 65 °C 72 °C	10 s 30 s 30 s	10
Final Extension	72 °C	2 min	1
Hold	4 °C	∞	

4.4.4.2 PCR 2

1. PCR round 2 is performed to attach specific Illumina TruSeq indexes i7 and i5 (*see* **Note 14)**. Set up 4 PCR reactions, one for each corresponding template DNA from the previous PCR reaction.

5X Q5 Reaction Buffer	10 μL
10 mM dNTPs	1 μL
10 μM PCR.3	2.5 μL
10 μM PCR.4	2.5 μL
Template DNA	x μL
Q5-HF DNA Polymerase	0.5 μL
H_2O	33.5-x μL
Total	**50 μL**

2. Template DNA indicates either primers only, AtRNL– T5Exo+ Tpt1+, AtRNL– T5Exo+ Tpt1+, or AtRNL+ T5Exo+ Tpt1+ product from the previous PCR reactions and only 5 μL for each reaction.

3. Run PCR with the following settings:

Cycle Step	**Temperature**	**Time**	**Cycles**
Initial Denaturation	98 °C	30 s	1
Denaturation Annealing Extension	98 °C 65 °C 72 °C	10 s 30 s 30 s	7-15
Final Extension	72 °C	2 min	1
Hold	4 °C	∞	

4.4.5 PAGE

1. Prepare loading samples, as follows:

   [DNA Ladder]

6X Sucrose Loading Buffer	2 μL
Quick-Load 50 bp DNA Ladder	2 μL
H_2O	8 μL
Total	**12 μL**

   [PCR Products]

6X Sucrose Loading Buffer	2 μL
PCR Product	10 μL
Total	**12 μL**

2. The samples are loaded on 6% Non-denaturing polyacrylamide gel with Quick-Load 50 bp DNA Ladder (New England Biolabs) as the ladder.
3. Stain the gel in 1X SYBR Gold for 40-45 min.
4. Visualize under UV light.
5. An exemplary gel image is shown in Fig. 5. AtRNL+ T5Exo+ Tpt1+ sample will be library, while Primers-only; AtRNL– T5Exo+ Tpt1+, and AtRNL+ T5Exo– Tpt1+ samples will be the controls.

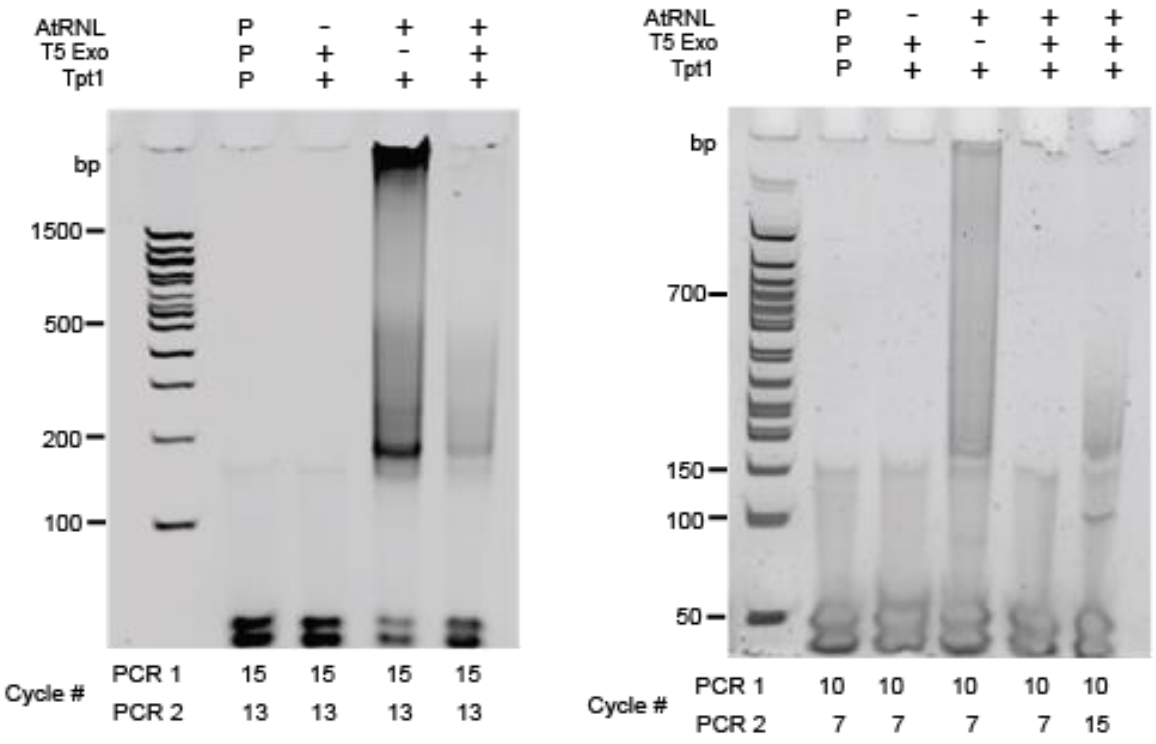


Figure 4.5 An example of ribose-seq library from yeast genomic DNA of *S. cerevisiae* cells. (Left) PCR products were analyzed by PAGE. Lane 1, 100 bp DNA Ladder; lane 2: control with PCR primers only (P); lane 3: control sample AtRNL-, T5 Exo+ and Tpt1+; lane 4: control sample AtRNL+, T5 Exo- and Tpt1+; lane 5: sample of ribose-seq library AtRNL+, T5 Exo+ and Tpt1+. PCR cycle numbers for PCR round 1 and PCR round 2 are as indicated at the bottom of the gel image. The smear seen in lane 5 shows the presence of ribose-seq library. (Right) PCR products are analyzed by PAGE. Lane 1, Quick-Load 50 bp DNA Ladder; lane 2: control with PCR primers only (P); lane 3: control sample AtRNL-, T5 Exo+ and Tpt1+; lane 4: control sample AtRNL+, T5 Exo- and Tpt1+; lane 5: sample of ribose-seq library AtRNL+, T5 Exo+ and Tpt1+ performed with low PCR cycles; lane 6: sample ribose-seq library AtRNL+, T5 Exo+ and Tpt1+ performed with high PCR cycles to confirm the presence of ribose-seq library. PCR cycle numbers for PCR round 1 and PCR round 2 are as indicated at the bottom of the gel image. The smear seen in lane 6 confirms the presence of ribose-seq library.

4.4.6 Size selection and gel purification

1. Perform multiple PCR reactions with the template DNA (AtRNL+ T5Exo– Tpt1+) to obtain sufficient ribose-seq library and load them on a non-denaturing polyacrylamide gel.
2. Use a sterile razor blade to cut out a slice of gel from all of the lanes corresponding to 200-700 bp. Transfer the gel piece into a 3-mL syringe to break down into smaller pieces in a 15 mL conical tube.
3. Add 3 mL of 0.1 M NaCl solution into the 15 mL conical tube and place it onto a rotator to incubate at 4° C overnight.
4. Filter the gel-submerged NaCl solution using a syringe with a 0.2-micron filter and perform ethanol precipitation.
5. Use Qubit 3.0 (dsDNA HS Assay Kit) to quantify the ribose-seq library.

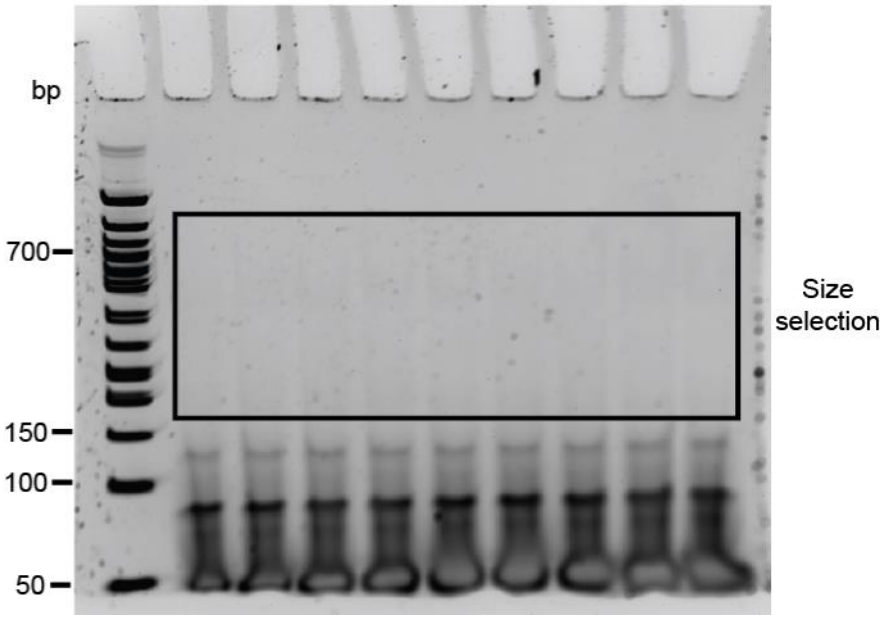


Figure 4.6 Size selection and gel purification of a yeast ribose-seq library before sequencing. Lane 1: Quick-Load 50 bp DNA Ladder; lanes 2-9: multiple samples of PCR round 2 with 7 cycles are run on a polyacrylamide gel for size selection and gel purification of a yeast ribose-seq library. DNA fragments between 200 bp and 700 bp are chosen (P: Primers-only, black-boxed region: selected area to be cut to extract and purify the ribose-seq library of interest).

4.5 Notes

1. Other yeast media and growth conditions can also be used, as far as the culture contains ~ 10^{10} cells. We typically obtain about 150-300 μg of genomic DNA, although ~40 is μg needed.
2. "L" and "S" stand for "Long" and "Short", respectively. "N" represents nucleotides of any base (A, C, G or T), as part of a random sequence.
3. Always use freshly prepared 2 M NaOH solution.

4. Bold letters indicate the specific index. In this example, we use TruSeq CD Index 701.
5. Bold letters indicate the specific index. In this example, we use TruSeq CD Index 501.
6. The amount of Zymolase can change depending on the volume of cell culture. Zymolase breaks down the cell wall to result in spheroblasts.
7. Equilibrate the Qiagen Genomic-tip 500/G with 10 mL of Buffer QBT before use. Vortex the supernatant before adding to the tip column. Additionally, pre-warm Buffer QF to 50 °C.
8. Pre-warmed Buffer QF can increase the yield of genomic DNA extraction.
9. The tubes are incubated at -80 °C to ensure effective precipitation.
10. In our laboratory, we typically obtain about 150-300 μ g of genomic DNA.
11. 5-fold excess amount of Adaptor.S is added to ensure that all of the Adaptor.L molecules are annealed to Adaptor.S. The remaining single-stranded Adaptor.S molecules are removed in the subsequent purification step.
12. Genomic DNA could be digested with a different set of restriction endonucleases that create blunt-ended fragments. Select those whose sites are well distributed in the genome and whose digestion results in fragments smaller than approximately 450 bp. In addition, it is recommended to use at least two sets of restriction enzymes to build two independent ribose-seq libraries for the same DNA samples to ensure that results obtained are independent from the restriction enzymes used.

13. Use a QIAGEN spin column for each reaction not to exceed the maximum binding capacity of the column.
14. In our example, we use TruSeq Indexes D701 and D501 to prepare the library. Alternatively, other Illumina TruSeq Indexes could be used.

4.6 Acknowledgements

We thank A. V. Bryksin, A. L. Gombolay, S. Biliya, and F. O. Vannberg for technical advises, and all of the Storici lab members for discussions and suggestions during the course of this project. This work was supported by the National Institutes of Health (R01ES026243-01 to F.S.), the Parker H. Petit Institute for Bioengineering and Bioscience at the Georgia Institute of Technology (12456H2 to F.S.), and the Howard Hughes Medical Institute Faculty Scholar grant (55108574 to F.S.).

CHAPTER 5. STUDYING RIBONUCLEOTIDE SPECTRUM IN GENOMIC DNA OF YEAST CELLS

The work in this Chapter is part of a research article in preparation for submission.

Balachander, S.[1†], Gombolay, A.[1†], Yang, T.[1], Xu, P.[1], Newnam, G.[1], El-Sayed, W.M.M.[1], Keskin, H,[1] Koh, K.D.[2], & Storici, F.[1]

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA.

[2]Department of Medicine, University of California, San Francisco, California, USA.

[†]Equal contribution

5.1 Summary

The cellular pools of ribonucleotides (rNTPs) are much higher than deoxyribonucleotides (dNTPs). Ribonucleoside monophosphates (rNMPs) are incorporated during DNA replication and repair of DNA damage. The presence of a highly reactive 2'-hydroxyl group in the ribose sugar alters the DNA properties, structure and function, thereby leading to genomic instability. We developed a technique called ribose-seq, described in Chapter 2, which enables us to make libraries of rNMP sites by directly capturing the rNMPs embedded in DNA. We modified our original method to improve the efficiency of ribose-seq technique over a factor of 1,000. In addition to mapping rNMPs in RNase H2 defective cells, we can now capture rNMPs in wild type (WT) RNase H cells. Overall, through multiple analyses, we found differences in the genomic composition of rNMPs in WT and RNase H mutant cells. At the site of incorporation, rCs and rGs were incorporated more frequently than expected, and low incorporation of rAs and particularly rUs were observed. Additionally, we have also obtained rNMP spectra in an RNase H2 mutant that lacks the interaction domain with the proliferation cell nuclear antigen (PCNA), and in a mutant, that impairs RNase H2 function at single rNMPs in DNA but not at long RNA-DNA hybrids. We found that removal of rNMPs in yeast does not require binding to PCNA, and rNMPs are primarily present as single nucleotides. From ribose-seq analysis of different strains of *S. cerevisiae,* and different yeast species including *S. paradoxus* and *S. pombe*, we found that presence or absence of RNase H2 does not affect the pattern of rNMP incorporation in yeast mitochondrial genome, whereas it impacted the rNMP profile in the

nuclear genome of these yeast species. Shown here are specific frequencies of rNMPs and a few examples of hotspots of rNMP incorporation in mitochondrial and nuclear genomes of WT and RNase H2 mutant cells.

5.2 Introduction

Modifications of nucleotides in DNA pose a major threat to the genomic integrity of the cells. rNMPs are considered to be the most common type of DNA lesion [6]. There are several sources of rNMP incorporation in the genome: (i) many DNA polymerases can incorporate rNMPs into genomic DNA, including human replicative DNA polymerase Pol δ [18], human mitochondrial Pol γ [75], yeast replicative Pol α, β and ε [12], and *E. coli* Pol V [16], (ii) incomplete elimination of RNA primers during Okazaki fragment maturation, and (iii) oxidative damage of deoxyribose resulting in formation of rNMPs [24]. So far, there are two known positive effects of rNMPs in DNA. The presence of two rNMPs in the mating type locus stalls the leading strand synthesis and helps initiate the replication-coupled recombination event, thus leading to mating type switching in *S. pombe* [42]. Secondly, the presence of rNMPs in DNA act as strand discrimination signals during lagging-strand synthesis by evoking a mismatch repair [170]. A lot is known about the negative consequences of rNMPs in DNA. The presence of rNMPs in DNA is deleterious as it leads to breaks, mutations and genome instability [68-71]. rNMPs embedded in DNA can alter its structural and mechanical properties [37].

rNMPs are efficiently removed from DNA by ribonucleotide excision repair (RER) [1]. RNase H2 of RER can cleave single and longer stretches of rNMPs in DNA. RNase H1 can target stretches of rNMPs in DNA containing four or more rNMPs.

Deletion of RNases H results in accumulation of rNMPs in yeast cells, and does not affect the survivability of cells, but deletion of either RNase H1 or H2 in mice is embryonic lethal. RNase H2-null murine embryonic fibroblasts (MEFs) activate a p53-dependent damage response [53]. In humans, hypomorphic mutations in any of the three subunits of RNase H2 are associated with the neurological syndrome of Aicardi- Goutieres (AGS) [54, 112].

Given the importance of RNase H2 in maintaining genome stability and its role in actively removing rNMPs, it is important to determine the location, identity and distribution of rNMPs in genomic DNA. Currently, there are four methods that can be used to map of rNMPs in DNA [68-71]. Among the four methods, ribose-seq, a technique that our lab developed in 2015 (discussed in Chapter 2), is the only method that can directly capture the rNMPs embedded in DNA, whereas the other methods locate nucleotides upstream or downstream of rNMP sites. Additionally, ribose-seq is specific to directly capturing rNMPs embedded in DNA and does not capture unprocessed Okazaki fragments, RNA primers, nicks or abasic sites in DNA.

Since then, a few studies have tried to understand the rNMP profile in genomic DNA. In order to understand if the changes in the levels of dNTPs affect frequency of rNMPs incorporated, Wanrooij *et al.* 2017 altered the ratio of dNTP/rNTP levels and observed an inverse relationship between the concentration of individual dNTPs and the amount of the corresponding rNMPs in mitochondria DNA, whereas in nuclear DNA the rNMP incorporation was only affected in the absence of RER [173]. Also, in the same study, they showed that the overall cellular dNTP pool dictates the relative frequencies of rNMPs embedded in mitochondrial DNA and are not affected by the presence or absence of RNase

H2. In human mitochondrial cells, it was shown that rNMPs were used as building blocks to maintain DNA replication when dNTP pools were limiting [174]. Despite these studies, a comprehensive genome wide mapping of rNMPs in nuclear and mitochondria of RER-proficient (active RNase H2) and RER-deficient cells is still missing.

In Chapter 4, we describe modifications made to the ribose-seq technique to capture the rNMPs more efficiently. Using the modified ribose-seq, we not only mapped rNMPs in RNase H2-null cells but also generated a profile of rNMPs in RNase H2 wild type (WT) cells. Using our ribose-seq analysis, we determined the spectrum of rNMPs in different yeast strains, and obtained the spectrum of rNMP incorporation in these cells. Additionally, we utilized several RNase H2-defective strains to compare the rNMP incorporation pattern between each other, and also compared these spectra with that obtained from WT RNase H cells. We observed that the rNMP incorporation pattern of the mitochondrial DNA genome is similar both in WT and RNase H-defective cells, thereby providing evidence that support absence of RNase H2 activity on rNMPs in mtDNA. Finally, we also compared the rNMP distribution in *S. cerevisiae*, *S. paradoxus* and *S. pombe.* All these results are helpful to better understanding the profile of rNMPs in genomic DNA.

5.3 Materials and methods

5.3.1 Construction of yeast strains

All the yeast strains used in this study are presented in **Table C.1**. We used haploid *Saccharomyces cerevisiae* strains from different background: E134, BY4742, and YFP17. In addition to *S. cerevisiae*, we also used *Saccharomyces paradoxus* (DG2204), and *Schizosaccharomyces pombe* (JZ105). All *RNH201* deletion strains were made by

replacing *RNH201* via transformation with a PCR product containing *hygMX4* or *kanMX4* cassette flanked by 50 nucleotides of sequence homologous to regions upstream and downstream of *RNH201* ORF. KK-172 was made from KK-44 by replacement of *RNH1* with the *kanMX4* cassette. Yeast strains SB-285 and SB-286 were derived from KK-44 by using the *delitto perfetto* method [175] and then by popping out the CORE cassette by a pair of oligonucleotides, 202PIP.F and 202PIP.R to mutate the proliferating cell nuclear antigen (PCNA) interacting peptide box, which is present in Rnh202 to make *rnh202*-FF346,347AA[44, 132]. SB-311 was derived from KK-2 by using the *delitto perfetto [175]* method to generate *rnh201*-P45D and *rnh201*-Y219A by using oligonucleotides, RNH P45D.60 and RNH Y219A.60 to result in a RNase H2 ribonucleotide excision defective (RED) [132] mutant. All mutations were confirmed by sequence analysis of PCR products obtained from amplification of a DNA region surrounding the specific mutation.

5.3.2 Ribose-seq library preparation

Ribose-seq libraries were prepared as previously described with some modifications (REF). Specifically, 40 μg of yeast genomic DNA were fragmented using restriction enzymes to produce blunt-ended fragments with an average size of 450 base pairs (bp) in length. Multiple sets of restriction enzymes were used for different library preparation, as shown in **Table C.2**. The different combinations used were (i) RsaI and HaeIII, (ii) DraI, EcoRV and SspI and (iii) DraI, EcoRV, SspI and AluI. Following restriction digestion, the fragmented DNA was purified by spin column (Qiagen). The fragments were tailed with dATP (Sigma-Aldrich) by using exo$^-$ Klenow fragment (NEB) for 30 min at 37 °C and purified by using spin column. Following dA-tailing and purification, the DNA fragments were annealed with a double-stranded adaptor (Adaptor.L1 or Adaptor.L2 with Adaptor.S,

Table C.3) by using T4 DNA ligase (NEB) incubating overnight at 15 °C. Following overnight ligation, the products were purified using AMPure XP beads (Beckman Coulter). The annealed fragments were treated with 0.2 M NaOH for 2 h at 55 °C to denature the DNA strands, and to cleave at the rNMP sites resulting in 2'-3'-cyclic phosphate and 2'-phosphate termini. This is followed with neutralization using 0.2 M HCl and purification using AMPure beads. All the successive purification steps were performed using AMPure XP Beads. The single-strand (ss) DNA fragments were incubated with 1 μM *Arabidopsis thaliana* tRNA ligase (AtRNL), 50 mM Tris-HCl pH 7.5, 40 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT and 30 μM ATP for 1 h at 30 °C, followed by purification. AtRNL aids in ligating the 2'-phosphate ends of rNMP-terminated ssDNA fragment to its opposite 5'-phosphate end, which results in a single-circular DNA. The fragments were then treated with T5 Exonuclease (NEB) for 1 h 30 min at 37 °C to degrade the unligated ssDNA fragments. After purification, the circular fragments were incubated with 1 μM 2'-phosphotransferase (Tpt1), 20 mM Tris-Hcl pH 7.5, 5 mM $MgCl_2$, 0.1 mM DTT, 0.4% Triton X-100 and 10 mM NAD^+ for 1 h at 30 °C to remove the 2'-phosphate present at the ligation junction. After Tpt1 treatment and purification, the circular fragments were PCR-amplified using two rounds of amplifications to result in ribose-seq library.

A first round of PCR was performed to amplify and introduce the sequences of Illumina TruSeq Index primers. The primers (PCR.1 and PCR.2) used for the first round were the same for all libraries. A second round of PCR was performed to attach specific indexes i7 and i5 for each library. The sequences of PCR primers and indexes can be found in **Table C.3**. PCR rounds 1 and 2 were performed using Q5-High Fidelity polymerase (NEB) for 10 and 7 cycles respectively (unless specified otherwise). Following the PCR cycles, the

ribose-seq library was loaded on a 6% non-denaturing polyacrylamide gel and stained using 1X SYBR Gold (Life Technologies) for 40-45 min. Fragments between 200-700 bp were cut and gel purified using the crush and soak method [176]. The resulting ribose-seq libraries were mixed at equimolar concentrations and normalized to 1.5 nM. The libraries were sequenced on an Illumina MiniSeq in the Molecular Evolution Core Facility at Georgia Tech.

5.3.3 Processing and alignment of sequencing reads.

The sequencing reads consist of an eight-nucleotide UMI, a three-nucleotide molecular barcode, the tagged nucleotide (the nucleotide tagged during ribose-seq from which the position of the embedded rNMP is determined), and the sequence directly downstream from the tagged nucleotide. The embedded rNMP and the sequence upstream from the rNMP is the reverse complement of the tagged nucleotide and the tagged nucleotide's downstream sequence, respectively. The UMI corresponds to cycles 1-6 and 10-11, the molecular barcode corresponds to cycles 7-9, the tagged nucleotide corresponds to cycle 12, and the tagged nucleotide's downstream sequence corresponds to cycles 13+ of the raw FASTQ sequences.

Before aligning the sequencing reads to the reference genome, the reads were trimmed based on sequencing quality and custom ribose-seq adaptor sequence using cutadapt 1.16 (-q 15 -a 'AGTTGCGACACGGATCTCTCA') [177]. In addition, reads shorter than 62 bases after trimming were discarded. Following quality control, the Alignment and Coordinate Modules of the Ribose-Map toolkit were used to process and analyze the sequencing data [172]. First, the Alignment Module filtered the trimmed reads by the

corresponding molecular barcode, aligned the filtered reads to the reference genome (sacCer2), and de-duplicated the aligned reads based on their UMI and aligned positions. Next, the Coordinate Module used the alignment results to determine the chromosomal coordinates of the embedded rNMPs and per nucleotide coverage of the embedded rNMPs. To allow comparison between sequencing libraries of different read depth, the per nucleotide coverage was calculated by normalizing rNMP counts to counts per hundred.

5.3.4 rNMP composition in the genome

The fasta files generated from sequencing the ribose-seq libraries are extracted using bedtool getfasta (bedtools v2.27.0). Each rNMP is counted separately, then divided by the total number of rNMPs incorporated in that corresponding library to obtain the rNMP percentage.

5.3.5 DNA sequence context of embedded rNMPs

Using the Sequence Module of Ribose-Map, the frequencies of the nucleotides at the sites of embedded rNMPs and 100 nucleotides upstream and downstream from those sites were calculated for both the nuclear and mitochondrial genomes. The nucleotide frequencies were then normalized to the frequencies of the corresponding region of the sacCer2 reference genome.

5.3.6 Data presentations

Graphs and plots were made using GraphPad Prism 5 (GraphPad Software).

5.4 Results

5.4.1 rNMP incorporation in WT vs RNase H2-null cells of *S. cerevisiae*

After performing the modified ribose-seq on WT and *rnh201*Δ cells, sequencing and analyzing the data, we first examined the absolute frequencies of rNMPs in the genome, and observed a noticeable difference in the rNMPs embedded in nuclear and mitochondria DNA. The absolute composition of genomic rNMPs were also different between WT and *rnh201*Δ cells (Table 5.1).

Table 5.1 Absolute rNMP composition in the genomic DNA of WT and *rnh201*Δ cells.

	WT		*rnh201*Δ	
	nDNA	mtDNA	nDNA	mtDNA
rA	26.75%	27.88%	21.14%	33.86%
rC	29.79%	24.60%	46.58%	34.89%
rG	11.07%	31.89%	24.99%	27.52%
rU	32.39%	15.63%	7.29%	3.73%

Shown in the table are the percentages of each rNMP quantified from ribose-seq libraries of WT (FS106) and *rnh201*Δ (FS138) of *S. cerevisiae* (BY4742 background). nDNA and mtDNA stand for nuclear DNA and mitochondrial DNA, respectively.

On normalizing the rNMP frequencies by the base composition of dNMPs of sacCer2 genome (**Table 5.2**), we observed a variation in relative frequencies of rNMPs in mitochondria DNA and nuclear DNA. In WT cells, rCMP levels are higher in nuclear DNA, whereas rGMP was highest in mitochondria DNA, followed by rCMP (**Figure 5.1**). In *rnh201*Δ, rC is most abundant, followed by rG and rA, and rU is the least incorporated

in nuclear DNA of *rnh201*Δ cells. On comparing rNMP frequencies in mitochondrial DNA between WT and *rnh201*Δ, there is a slight increase in rC and rA, slight decrease in rG but very low incorporation of rU in mitochondrial DNA of *rnh201*Δ cells (**Figure 5.1**).

Table 5.2 Base composition of genomic DNA in *S. cerevisiae, S. paradoxus,* and *S. pombe.*

	S. cerevisiae		*S. paradoxus*		*S. pombe*	
	Nuclear	Mitochondria	Nuclear	Mitochondria	Nuclear	Mitochondria
A	30.85%	41.45%	30.81%	42.35%	31.97%	34.96%
C	19.15%	8.55%	19.19%	7.65%	18.03%	15.04%
G	19.15%	8.55%	19.19%	7.65%	18.03%	15.04%
T	30.85%	41.45%	30.81%	42.35%	31.97%	34.96%

Shown above are the individual percentages of dNMPs (A, C, G and T) in nuclear and mitochondria genome of *S. cerevisiae* (sacCer2), *S. paradoxus* (NCBI database), and *S. pombe* (Pombase).

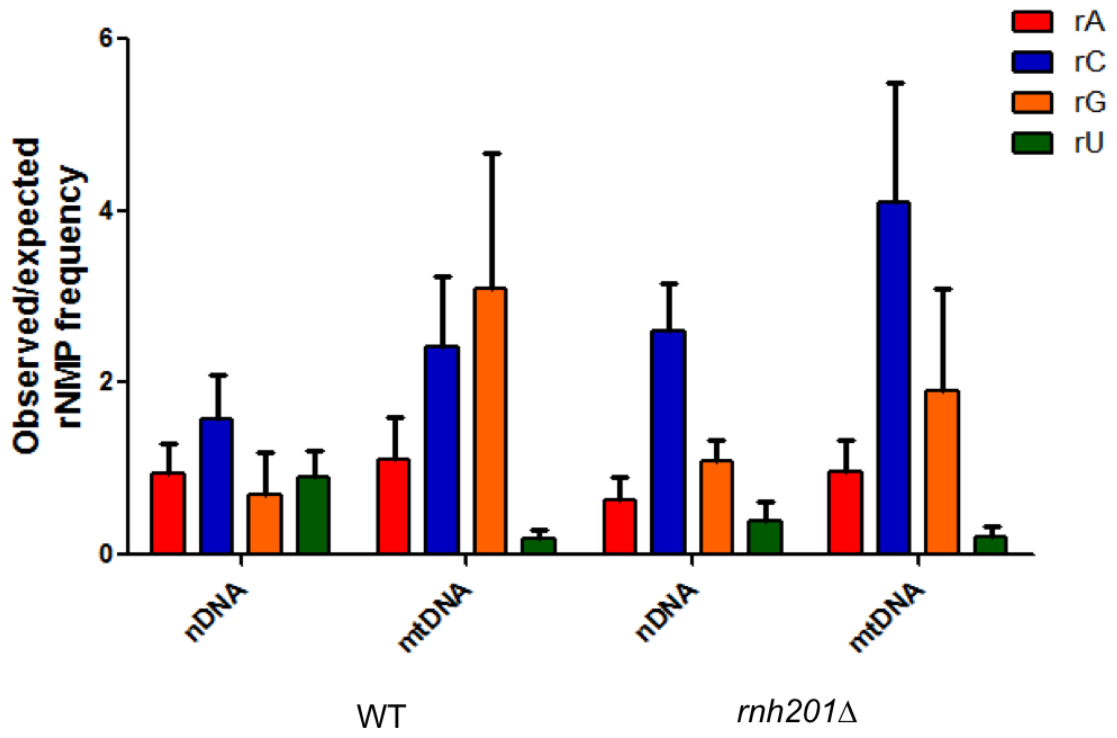


Figure 5.1 Normalized ratios of rNMP frequency in genome of *S. cerevisiae.* The observed frequencies of rNMPs from **Table 5.2** were divided by the expected frequencies of each base in nuclear DNA (nDNA) and mitochondrial DNA (mtDNA) of sacCer2 genome (**Table 5.1**). Shown here are bar graphs representing mean with standard deviation for WT (n=7) and *rnh201*Δ (n=10).

Next, we studied the DNA sequence context at the site of rNMP embedded in DNA. This helps us to identify specific sites of rNMP incorporation and to characterize the surrounding DNA sequence context (100 nucleotides upstream and downstream) (**Figure 5.2**). On comparing the plots generated from WT and *rnh201*Δ ribose-seq libraries, we observe the following: (i) a difference in spectrum of rNMP in nuclear genome between WT and *rnh201*Δ. There is no bias toward a specific rNMP in the nuclear DNA of WT as

the frequencies of nucleotides are similar to the background frequencies. In contrast, we observe rCMP and rUMP as most and least frequently incorporated rNMP, respectively, in the nuclear DNA of *rnh201*Δ cells.

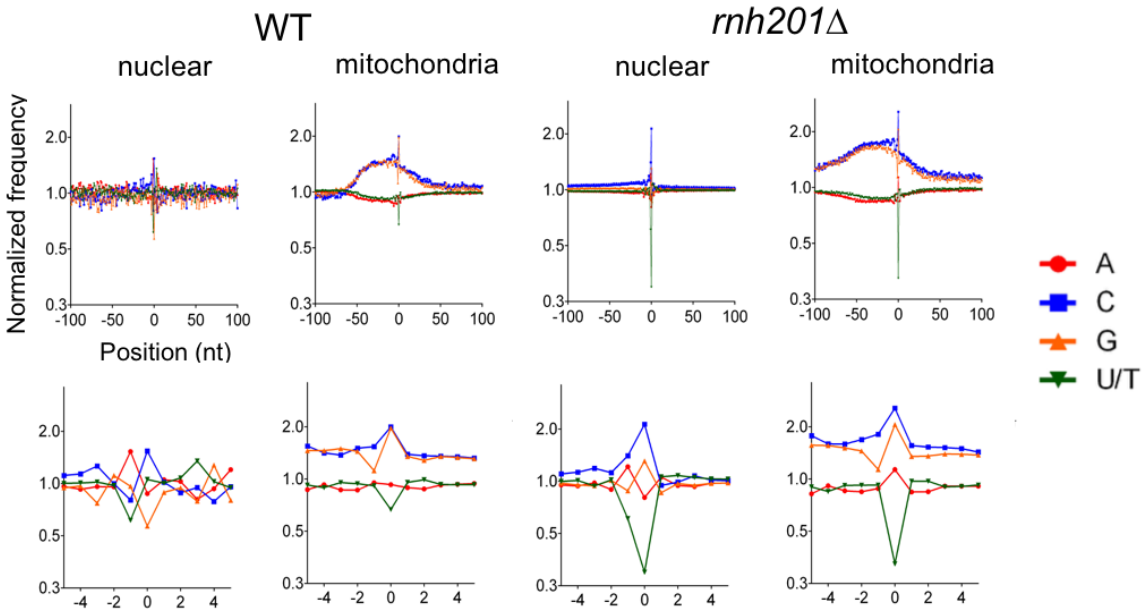


Figure 5.2 Identity and sequence context of rNMP incorporation in genomic DNA of *S. cerevisiae*. Shown here are the normalized nucleotide frequencies relative to mapped positions of sequences from ribose-seq library of WT (left side two columns) and *rnh201*Δ (right side two columns). Position 0 is the rNMP. + and – nucleotides are downstream and upstream dNMPs, respectively. Zoom-out and zoom-in of nucleotide frequencies are shown in the top and bottom row, respectively. *S. cerevisiae* strain used were KK-2 (WT, BY4742) and SB-285 (*rnh201*Δ).

(ii) Similar rNMP incorporation pattern in mitochondrial genome of WT and *rnh201*Δ, indicating that RNase H2 does not have a marked role in removing rNMPs in mitochondrial DNA. This is consistent with a recent study that demonstrated that yeast mitochondria lack an efficient mechanism (RER) for removing rNMPs that are misincorporated by mitochondrial polymerase during replication [173, 174]. (iii) Although, mitochondria

being rich in A+T content (82.9%), the site of rNMP incorporation was within or downstream of G+C rich regions (**Figure 5.2**). The vast majority of G and C nucleotides, despite being only 17.1%, in the mitochondrial genome of *S. cerevisiae,* are clustered together [178]. It could be possible that these rNMPs in the G+C rich regions can help trigger DNA recombination, which can in turn initiate DNA replication in mitochondria of yeast [102].

Comparing the overall rNMP frequencies in nuclear DNA of WT and *rnh201*Δ, we noticed an increase in rCMP and rGMP frequencies in *rnh201*Δ cells, and a strong decrease in frequency of rUMP (**Table 5.1, Figures 5.1** and **5.2**). Despite the difference in absolute percentages of rG, rC and rA in mitochondria DNA (**Table 5.1**), we found that rCs and rGs were incorporated more frequently than expected, hence at the site of rNMP incorporation we found higher frequency of rC or rG incorporation than of rA (**Figure 5.2**). rUMPs were found to be the least incorporated in nuclear and mitochondrial genome. We previously showed that rUMPs are not removed by uracil-N-glycosylase 1 (UNG1), as shown in Chapter 1. Low frequency of rUMP in genomic DNA may be due to the high concentrations of dTTPs in the nucleotide pool. dCTP and dGTP are the least abundant dNTPs, therefore they might be depleted faster than dTTP and dATP, thereby lowering the incorporation rate of rUMP [18].

5.4.2 Fragmentation of genomic DNA using different sets of restriction enzymes

A crucial step in the preparation of a ribose-seq library is the fragmentation of genomic DNA using blunt-ended restriction enzymes. To test if the restriction enzymes used resulted in any bias to the rNMP incorporation, we used three different sets of restriction

enzymes (**Table C.2**). Comparing rNMP incorporation in *rnh201*Δ cells when fragmented using different restriction enzymes: DraI + EcoRV + SspI, RsaI + HaeIII, or DraI + EcoRV + SspI + AluI, we observed no major differences in the spectrum of rNMPs in both nuclear and mitochondrial genome (**Figure 5.3)**. On using RsaI + HaeIII, we saw a slight variation in mitochondria plots. This is likely due to the sequence recognition of HaeIII and RsaII as compared to the other sets, but the overall pattern of rNMP incorporation stayed the same. Additionally, we analyzed four independent libraries for *rnh201*Δ strains and found that replicate libraries yielded similar results (**Figure C.1**).

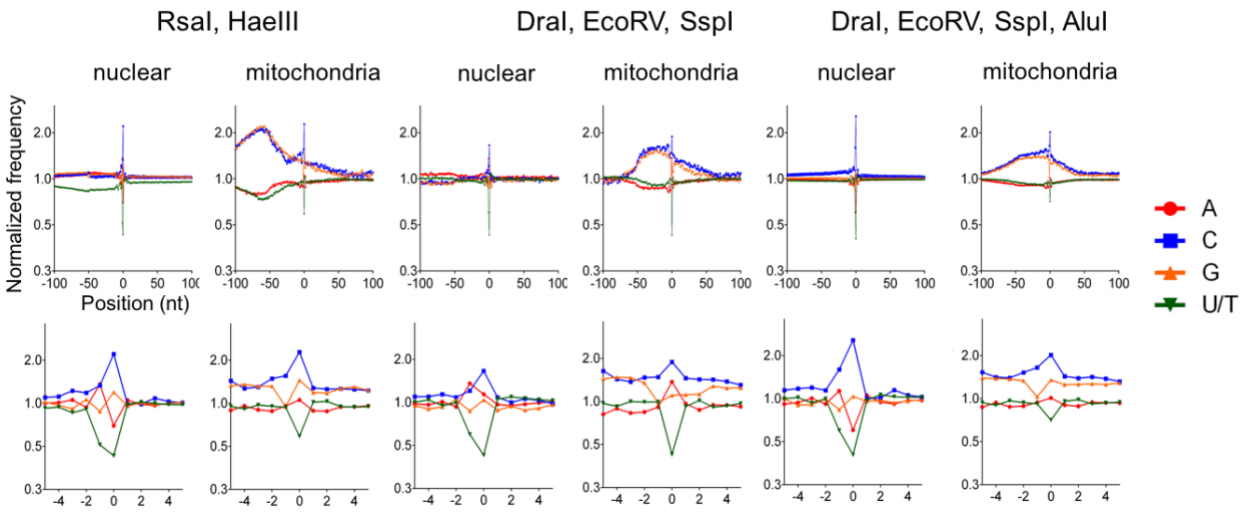


Figure 5.3 Identity and sequence context of rNMP incorporation in *S. cerevisiae* *rnh201*Δ genome when using different restriction enzymes for fragmentation of genomic DNA. Shown here are normalized nucleotide frequencies relative to mapped positions of sequences from ribose-seq library *rnh201*Δ. Position 0 is the rNMP. + and – nucleotides are downstream and upstream dNMPs, respectively. Zoom-out and zoom-in of nucleotide frequencies are shown in the top and bottom row, respectively.

5.4.3 Comparing rNMP incorporation in different *S. cerevisiae* background

To examine if the rNMP incorporation that we observed were consistent in different *S. cerevisiae* strains, we prepared ribose-seq libraries from additional *S. cerevisiae* strains. Shown in **Figure 5.4** and **5.5** are rNMP incorporation pattern obtained from YFP17 and E134. The pattern of rNMP incorporation from these backgrounds were similar to that observed when using BY4742 strain (**Figure 5.2**), both WT and *rnh201***Δ** cells. After normalization of of rNMP frequencies by the base composition, we observed that rC and rG were frequently incorporated, while rA and particularly rU were the least incorporated rNMP (**Figures 5.4 and 5.5**).

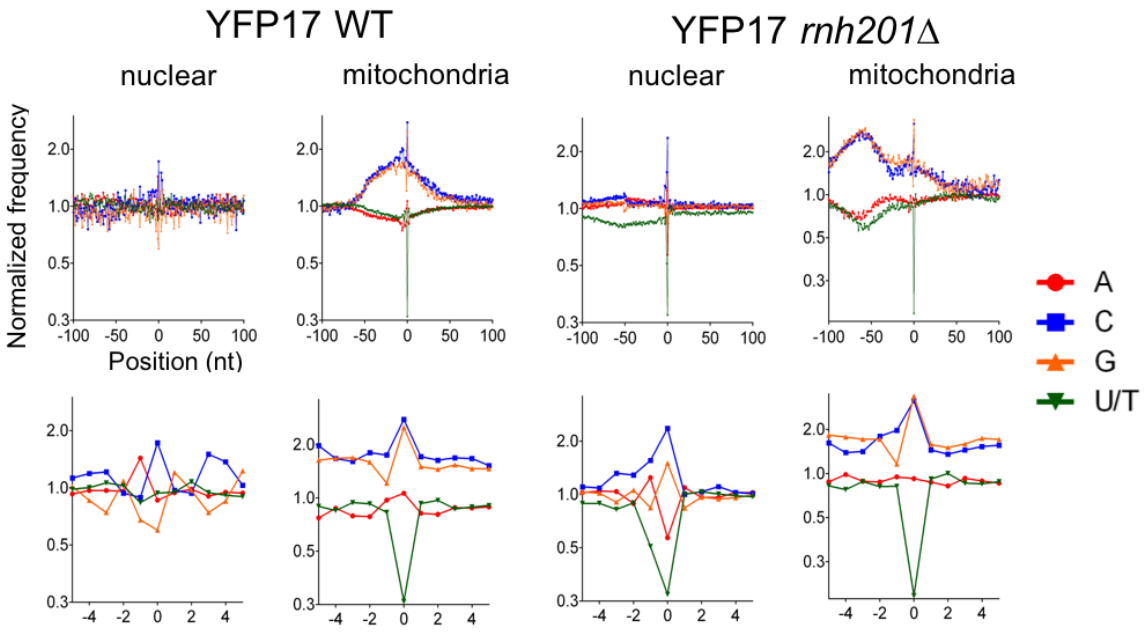


Figure 5.4 Identity and sequence context of rNMP incorporation in *S. cerevisiae* YFP17 genome. Shown here are normalized nucleotide frequencies relative to mapped positions of sequences from ribose-seq library of WT (left side two columns) and *rnh201*Δ (right side two columns). Position 0 is the rNMP. + and – nucleotides are downstream and upstream dNMPs, respectively. Zoom-out and zoom-in of nucleotide frequencies are shown in the top and bottom row, respectively.

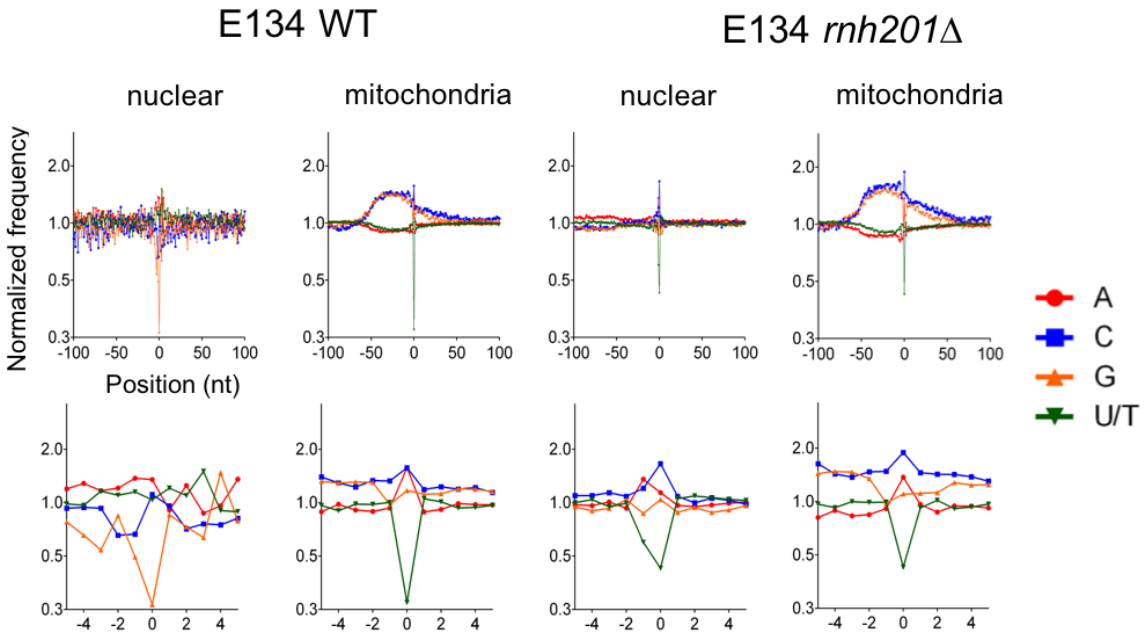


Figure 5.5 Identity and sequence context of rNMP incorporation in *S. cerevisiae* E134 genome. Shown here are normalized nucleotide frequencies relative to mapped positions of sequences from ribose-seq library of WT (left side two columns) and *rnh201*Δ (right side two columns). Position 0 is the rNMP. + and – nucleotides are downstream and upstream dNMPs, respectively. Zoom-out and zoom-in of nucleotide frequencies are shown in the top and bottom row, respectively.

5.4.4 rNMP profile in DNA of various RNase H2-mutants of *S. cerevisiae*

RNase H2 recognizes both single rNMP embedded in DNA as well as stretches of RNA in DNA. Several mutants of RNase H2 have been characterized to perform different functions. A separation of function mutant of RNase H2 (Rnh201-P45D,Y219A), called ribonucleotide excision defective (RED), was found to retain its activity on RNA-DNA hybrids but not on single rNMPs [132]. So, analyzing the rNMP spectrum of RNase H2-RED would give us a better understanding of the location preferences for single rNMPs in

DNA. In addition to a RED strain, we also performed ribose-seq on a mutant carrying mutation in the Rnh202 subunit of RNase H2. Studies *in vitro* have suggested an additional role of RNase H2 during DNA replication. Rnh202, subunit of RNase H2, has a PCNA-interacting peptide domain (PIP-box) which mediates the interaction of RNase H2 and PCNA, and localizes the RNase H2 to the replication fork [132]. Contrary to that, *in vitro* studies in human cells showed that RNase H2 can remove rNMPs in a PCNA-independent manner [46]. The absolute genomic composition of rNMPs are shown in **Table 5.3** for WT, *rnh202*-PIP, *rnh201*Δ and *rnh201*-RED strains.

Table 5.3 Absolute composition of genomic rNMPs in WT, PIP, *rnh201*Δ, and RED mutants.

	WT		PIP mutant		*rnh201*Δ		RED mutant	
	nDNA	mtDNA	nDNA	mtDNA	nDNA	mtDNA	nDNA	mtDNA
rA	26.75%	27.88%	31.03%	45.65%	21.14%	33.86%	24.07%	32.91%
rC	29.79%	24.60%	25.22%	32.37%	46.58%	34.89%	37.54%	21.53%
rG	11.07%	31.89%	15.52%	16.78%	24.99%	27.52%	30.32%	39.06%
rU	32.39%	15.63%	28.23%	5.30%	7.29%	3.73%	8.07%	6.50%

Shown above are the percentages of each dNMPs (A, C, G and T) estimated in nuclear and mitochondria DNA of WT, PIP (Rnh202-FF346,347AA), *rnh201*Δ, and RED (Rnh201-P45D,Y219A) mutants. nDNA and mtDNA stand for nuclear DNA and mitochondrial DNA, respectively.

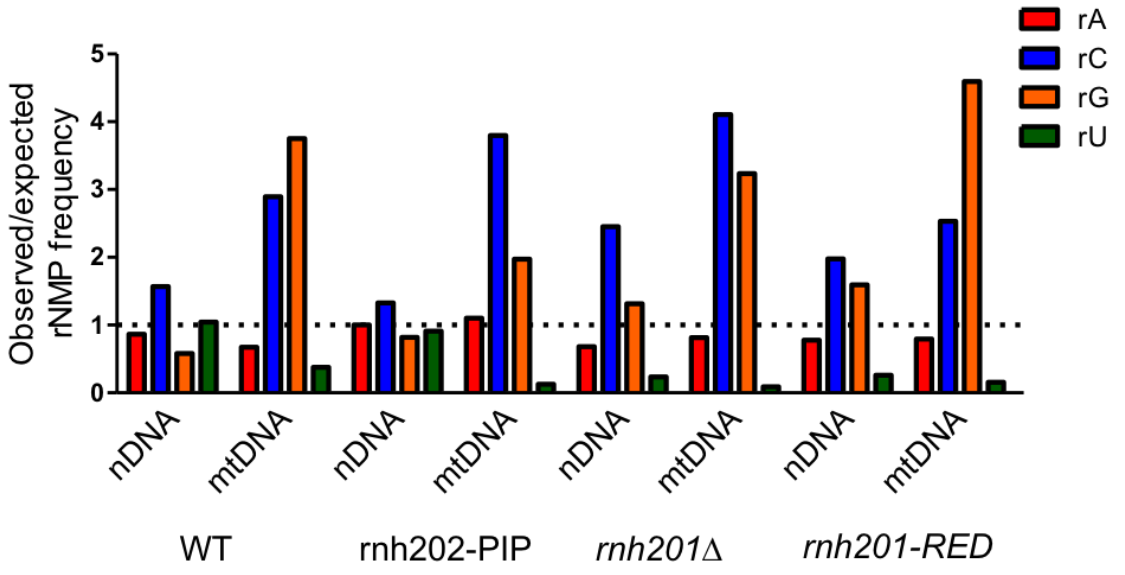


Figure 5.6 Normalized ratios of rNMP in WT, *rnh202*-FF346,347AA, *rnh201*Δ, and *rnh201*-P45D,Y219A cells. The observed rNMP frequencies (**Figure 5.1** and **5.6)** were divided by the expected proportion of bases shown in **Table 5.2.** Shown here are bar graphs from one library of each mutant.

On normalizing the relative frequencies of rNMPs, we observed that rG was the preferred rNMP in mitochondria DNA of RED and WT cells, whereas rC was the preferred in mitochondria DNA of PIP and *rnh201*Δ cells (**Figure 5.6**). On the contrary, in nuclear DNA of WT, PIP, *rnh201*Δ and RED strains, rC was more frequently incorporated (**Figure 5.6**).

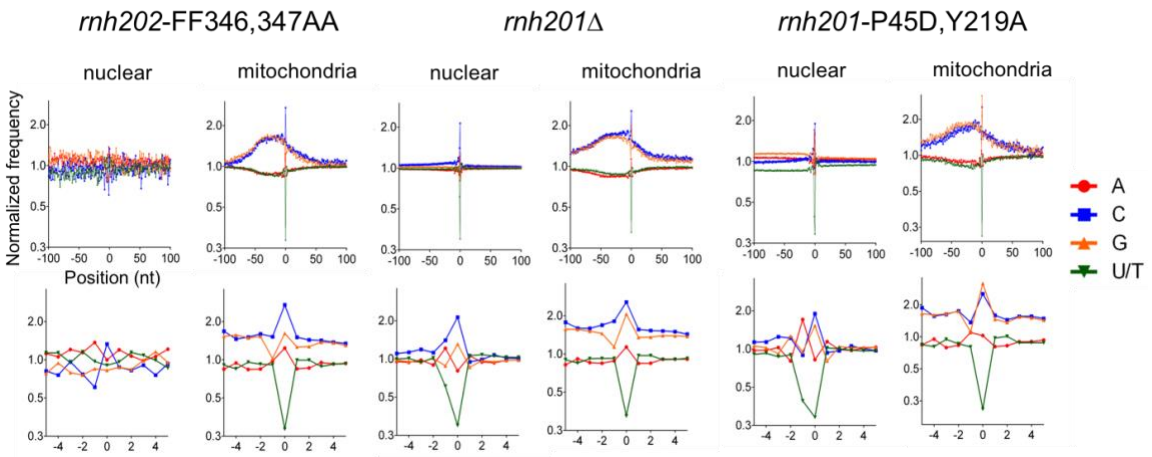


Figure 5.7 Profile of rNMP incorporation in RNase H2 mutants from *S. cerevisiae* BY4742 cells. Normalized nucleotide frequencies relative to mapped positions of sequences from ribose-seq library of WT (left side two columns) and *rnh201*Δ (right side two columns). Position 0 is the rNMP. + and – nucleotides are downstream and upstream dNMPs, respectively. Zoom-out and zoom-in of nucleotide frequencies are shown in the top and bottom row.

Overall, the spectra of rNMP incorporation in *rnh202*-PIP and WT cells were similar (**Figure 5.5**), suggesting that *in vivo* RNase H2 can remove rNMPs via a PCNA-independent manner. Interestingly, the spectra of rNMP incorporation in *rnh201*Δ and *rnh201*-RED cells were similar (**Figure 5.5**), which suggests that rNMPs may primarily be present in the form of single nucleotides embedded in DNA.

5.4.5 Pattern of rNMP incorporation in *S. paradoxus* and *S. pombe*

Having determined the rNMP profile in *S. cerevisiae*, we wanted to analyse the pattern of rNMP incorporation in *S. paradoxus* (budding yeast more similar to *S. cerevisiae*) and *S. pombe* (fission yeast). *S. cerevisiae* is evolutionarily closer to *S. paradoxus*, and more distant to *S. pombe* [179, 180]. The A+T and G+C content of *S. cerevisiae* and *S. paradoxus*

are similar, but the G+C content of *S. pombe*'s mitochondrial DNA is almost twice than that of budding yeast (**Table 5.2**).

On performing ribose-seq analysis, we quantified the absolute composition of genomic rNMPs (**Table 5.4**). As observed in *S. cerevisiae* (**Table 5.1** and **Figure 5.1**), there was a difference in pattern of rNMP incorporation between WT and *rnh201*Δ cells of *S. paradoxus* and *S. pombe*. Higher incorporation of rG and rC were seen in WT and *rnh201*Δ strains of *S. paradoxus*, respectively. In *S. pombe*, rC and rU were more frequently incorporated in WT nuclear DNA, whereas a higher incorporation of rA was observed in nuclear DNA of *rnh201*Δ. rC was abundant in WT and *rnh201*Δ mitochondria of *S. paradoxus*, whereas rG was incorporated frequently in WT and *rnh201*Δ mitochondria of *S. pombe*.

Table 5.4 Absolute composition of genomic rNMPs in *S. paradoxus* and *S. pombe.*

	S. paradoxus				*S. pombe*			
	WT		*rnh201*Δ		WT		*rnh201*Δ	
	nDNA	mtDNA	nDNA	mtDNA	nDNA	mtDNA	nDNA	mtDNA
rA	16.10%	25.58%	16.12%	33.66%	30.73%	24.88%	51.88%	31.45%
rC	22.06%	39.67%	49.25%	34.40%	24.43%	21.89%	18.86%	23.69%
rG	42.19%	30.90%	28.44%	29.03%	8.56%	47.47%	21.21%	31.25%
rU	19.65%	3.85%	6.19%	2.90%	36.27%	10.76%	8.05%	13.61%

Shown above are the absolute percentages of each rNMPs (A, C, G and U) in nuclear and mitochondria DNA of WT and *rnh201*Δ of *S. paradoxus* and *S. pombe*.

Upon normalizing the rNMP frequencies in nuclear and mitochondrial DNA, we found a relatively higher incorporation of rG and rC in *S. paradoxus*, which was similar to the pattern seen in *S. cerevisiae* (**Figures 5.1** and **5.8**). In *S. pombe*, we observed a higher frequency of rG incorporation in mitochondrial DNA of both WT and *rnh201*Δ cells, whereas rC and rA frequencies were higher in nuclear DNA of WT and *rnh201*Δ cells.

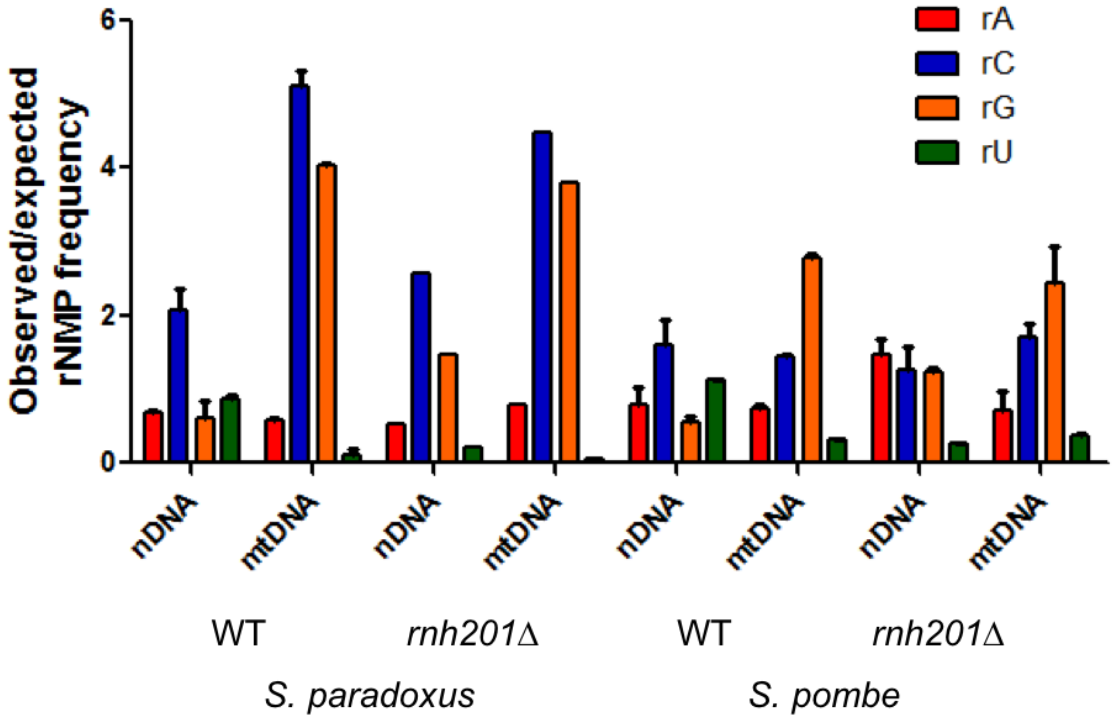


Figure 5.8 Normalized ratios of rNMP frequency in *S. paradoxus* and *S. pombe*. The observed frequencies from Table were divided by the proposition of each base in nuclear DNA (nDNA) and mitochondrial DNA (mtDNA). The absolute rNMP composition used are shown in **Table C.4**. Shown here are bar graphs representing mean with standard deviation for *S. paradoxus* WT (n=2) and *rnh201*Δ (n=1), and *S. pombe* WT (n=2) and *rnh201*Δ (n=2).

Next, we examined the site of rNMP incorporation and the dNMP sequence context at the site of rNMP incorporation. The spectra of rNMP incorporation in the mitochondrial DNA of WT and *rnh201*Δ cells in *S. paradoxus* were similar to that of *S. cerevisiae*, where

rNMPs were frequently located within or around the G+C rich regions, whereas this was not observed in *S. pombe* (**Figures 5.2, 5.4, 5.5, 5.9** and **5.10).** Although *S. pombe* had a higher mitochondrial G+C content, the G+C clustering is a unique feature of budding yeasts (**Figure 5.8)**.

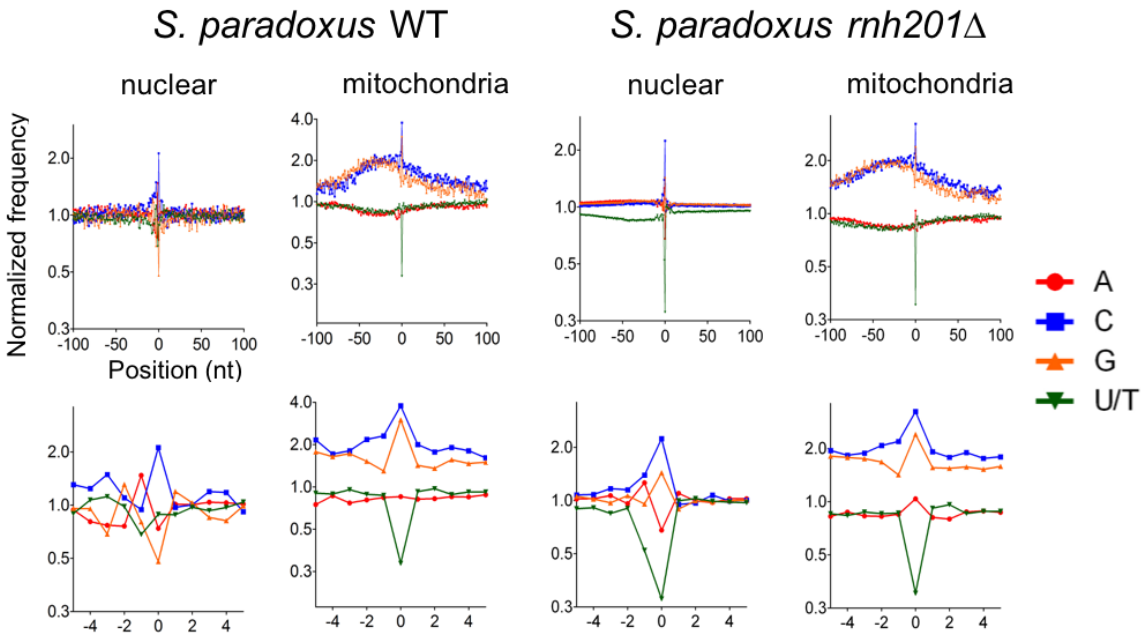


Figure 5.9 rNMP incorporation in WT and *rnh201*Δ cells of *S. paradoxus*. Normalized nucleotide frequencies relative to mapped positions of sequences from ribose-seq library of WT (left side two columns) and *rnh201*Δ (right side two columns). Position 0 is the rNMP. + and – nucleotides are downstream and upstream dNMPs, respectively. Zoom-out and zoom-in of nucleotide frequencies are shown in the top and bottom row.

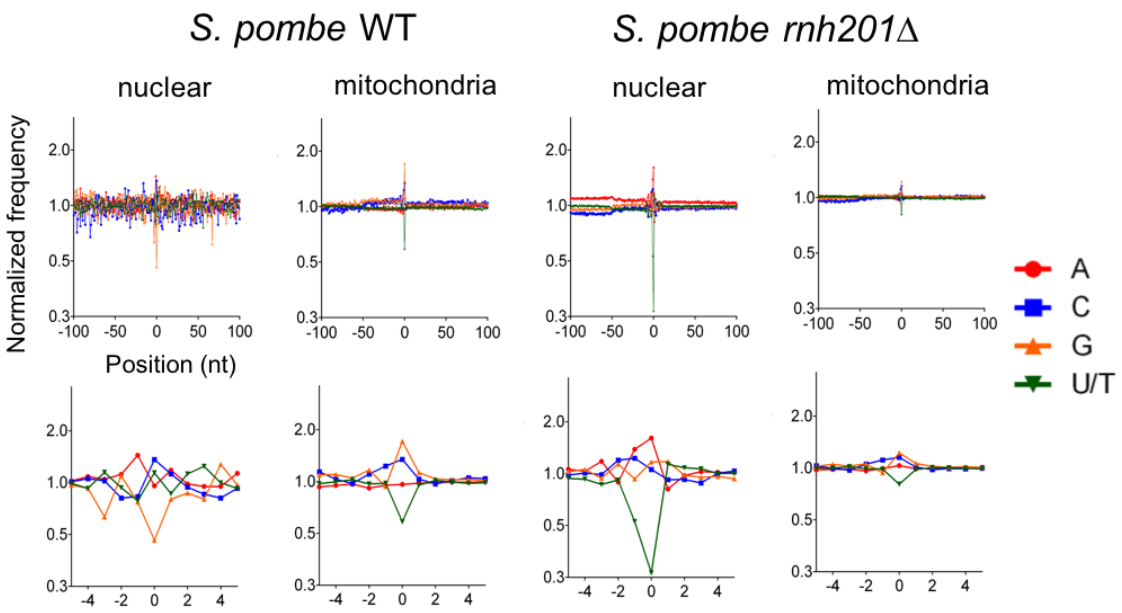


Figure 5.10 rNMP incorporation in WT and *rnh201*Δ cells of *S. pombe*. Normalized nucleotide frequencies relative to mapped positions of sequences from ribose-seq library of WT (left side two columns) and *rnh201Δ* (right side two columns). Position 0 is the rNMP. + and – nucleotides are downstream and upstream dNMPs, respectively. Zoom-out and zoom-in of nucleotide frequencies are shown in the top and bottom row.

5.5 Discussion

Despite the existence of at least four rNMP mapping techniques, the rNMP profile of cells having an active RER mechanism is completely missing, and the profile in RER defective cells has only been minimally characterized. Here, we used the optimized ribose-seq technique, as discussed in Chapter 4, to capture rNMPs in yeast wild-type RNase H2 cells and compare the rNMP spectrum of WT RNase H2 cells with that obtained in RNase H2-null and other RNase H2 mutant cells. First, we observed that by using different sets of restriction enzymes to fragment the genomic DNA, we could not detect major differences in the rNMP spectra in both nuclear and mitochondrial DNA of yeast *S. cerevisiae*. In

addition to confirm reproducibility of the results, the use of different sets of restriction enzymes was exploited to obtain a larger coverage of the genome with the preparation of the ribose-seq libraries. We found that the levels of rNMP incorporation were different between WT and *rnh201*Δ strains. The overall composition of genomic rNMPs varied depending on the presence or absence of RNase H2 in nuclear DNA, but had almost no impact on the mitochondrial DNA. By inspecting the identities of the rNMPs at the sites of incorporation in nuclear and mitochondrial DNA of *rnh201*Δ cells, we found that rC was more frequently incorporated than rG, whereas rA, and particularly rU were the least preferred. In WT cells, rC was preferred in nuclear DNA, whereas rG had higher incorporation frequency in mitochondrial DNA of WT cells. An overall summary of the highest and lowest rNMP incorporated in nuclear and mitochondrial genome is shown in **Table 5.5**.

Table 5.5 rNMP incorporation in nuclear and mitochondria DNA of WT and *rnh201*Δ cells of *S. cerevisiae, S. paradoxus*, and *S. pombe.*

	WT				RNase H2-null			
DNA	Nuclear		Mitochondria		Nuclear		Mitochondria	
	Highest	Lowest	Highest	Lowest	Highest	Lowest	Highest	Lowest
S. cerevisiae	A,C	G	C	U	C	U	C	U
S. paradoxus	C,G	G	C	U	C	U	C	U
S. pombe	C	G	G	U	A	U	C,G	U

Shown here are the identity of rNMP that are most and least (relative frequency) incorporated in WT and RNase H2-null of *S. cerevisiae*, *S. paradoxus*, and *S. pombe*. The rNMPs shown here are collective representation of normalized frequency plots that are shown in **Figures 5.2, 5.4, 5.5, 5.9 and 5.10**.

The rNMP incorporation frequencies may be attributed to the variation in the levels of corresponding dNTPs. dCTP and dGTP are the least abundant dNTPs, and hence might be depleted faster than dTTP and dATP, increasing the probability of rCMP and rGMP incorporation over rAMP and rUMP. Also, the low incorporation of rU could be due to the abundance of dTTPs in the nucleotide pool.

Interestingly, by performing ribose-seq on *rnh202*-PIP and *rnh201*-RED, we found that the profile of rNMPs from *rnh202*-PIP cells looked similar to that of WT cells, and the rNMP profile in *rnh201*-RED cells was similar to that obtained in *rnh201*Δ cells. The rNMP incorporation pattern in mitochondrial genome of WT and various RNase H2 mutant cells looked very similar to each other, thus providing additional evidence that support absence of RNase H2 activity on single rNMPs in mitochondria.

From ribose-seq analysis of different strains of *S. cerevisiae,* and different yeast species, *S. paradoxus* and *S. pombe*, we found that presence or absence of RNase H2 does not affect the pattern of rNMP incorporation in yeast mitochondrial genome, whereas it impacted the rNMP profile in the nuclear genome of these yeast species. We also observed similarity in rNMP profile between *S. cerevisiae* and *S. paradoxus*, but distinct rNMP profile in *S. pombe.* This was especially noticeable in the mitochondrial DNA. rNMPs were seen to be incorporated in G+C rich region of *S. cerevisiae* and *S. paradoxus*, and not in *S. pombe.*

In addition to studying the rNMP incorporation pattern, there are several other analyses that we are currently performing using our ribose-seq data. For example, we can look at the distribution and hotspots of rNMP incorporation throughout the genome. In **Figure**

5.11, we can observe the distribution of rNMPs in the mitochondrial genome of WT and *rnh201*Δ cells. Using our ribose-seq analysis, we can map the location of rNMPs in genomic DNA to look for hotspots. For example, we found hotspots of rNMPs downstream of the *COX2* gene that were present in WT and *rnh201*Δ cells of several *S. cerevisiae* strains (**Figure 5.12**).

Overall, our findings strongly suggest that rNMPs may not be randomly incorporated in the genome. Currently, we are analyzing if the sequence context has a role in determining the identify of rNMP incorporation adjacent to them. In addition to that, some of the questions we pose are; Can we derive specific rules for rNMP incorporation in yeast DNA? It would then be interesting to extend the application of our ribose-seq method to study rNMP incorporation also in higher eukaryotes. How does the rNMP spectrum change from yeast to human in nuclear and mitochondrial DNA? Does rNMP incorporation follow specific rules that are conserved across the evolutionary tree, or does every organism have their unique rNMP pattern?

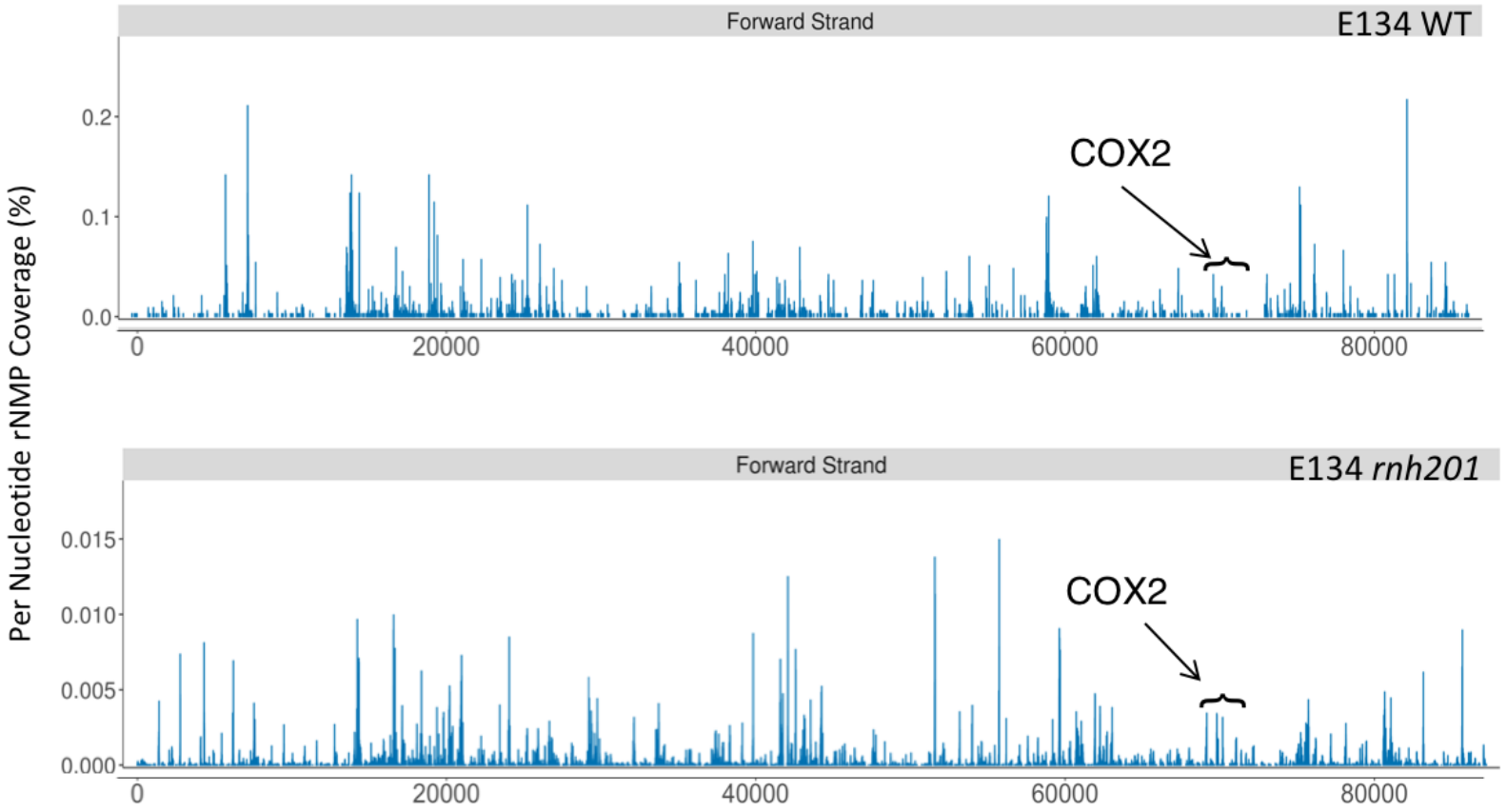


Figure 5.11 Distribution of rNMP incorporation in mitochondrial genome of *S. cerevisiae*. Ribose-seq map of rNMPs in mitochondrial DNA of WT (top) and *rnh201*Δ (bottom) showing persistent rNMPs in the genome of mitochondria. X-axis represents the coordinate of nucleotides in the mitochondrial DNA and y-axis is per nucleotide rNMP coverage and is calculated as (number of rNMPs in a specific spot/total rNMPs in the corresponding library) multiplied by 100.

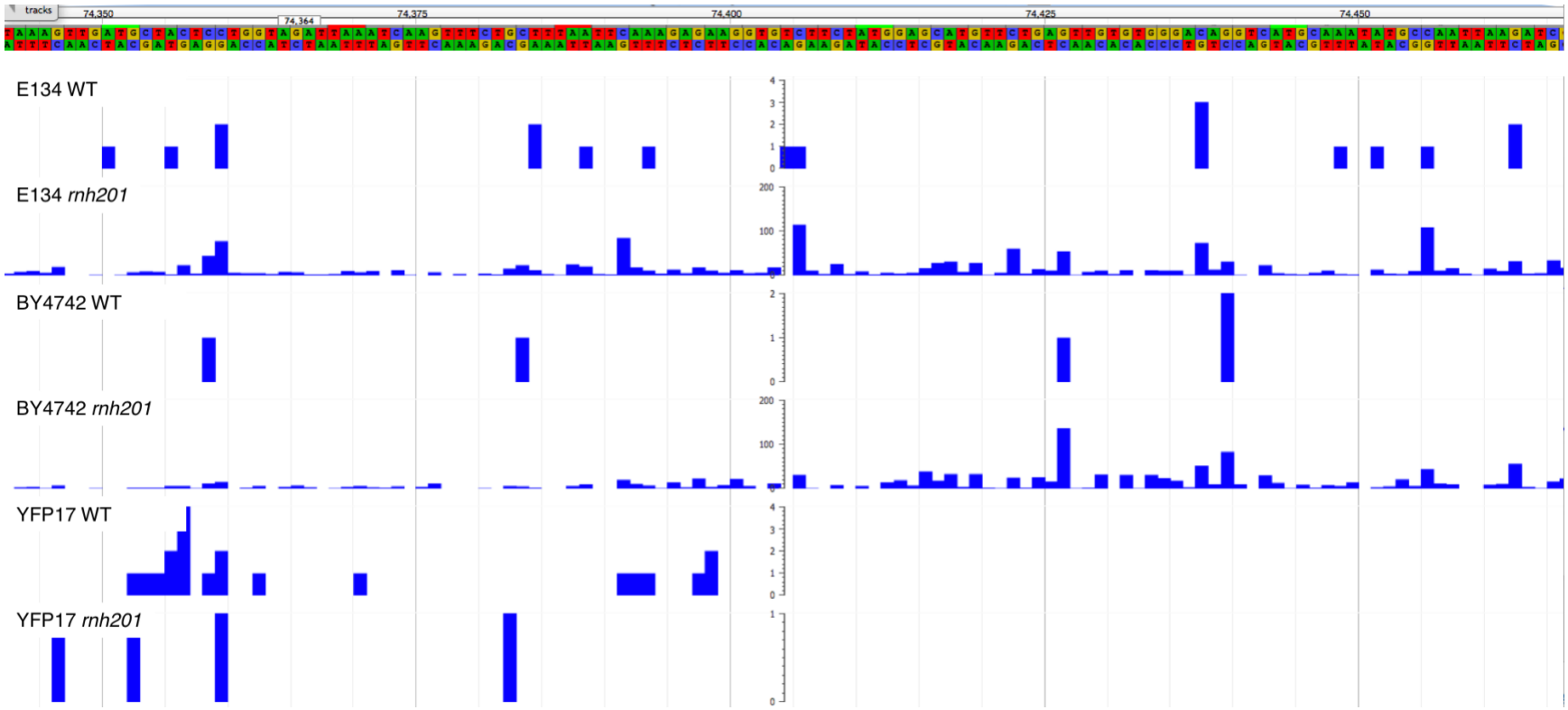


Figure 5.12 Hotspots of rNMP incorporation in *COX2* region of mitochondria DNA in *S. cerevisiae*. Genome browser view of rNMP peaks (hotspots) observed in several *S. cerevisiae* WT and *rnh201*Δ libraries.

5.6 Acknowledgements

We thank A. V. Bryksin, N. Djeddar, S. Biliya, and F. O. Vannberg for technical advice, and all of the Storici lab members for discussions and suggestions during the course of this project. This work was supported by the National Institutes of Health (R01ES026243-01 to F.S.), the Parker H. Petit Institute for Bioengineering and Bioscience at the Georgia Institute of Technology (12456H2 to F.S.), and the Howard Hughes Medical Institute Faculty Scholar grant (55108574 to F.S.).

CHAPTER 6. CONCLUSIONS AND FUTURE PERSPECTIVE

6.1 Overall conclusions

- Optimized the ribose-seq protocol to enhance the capability of capturing rNMPs by a factor of 1000, and using this modified method we map the profile of rNMPs in WT and RNase H2-null cells of *S. cerevisiae, S. paradoxus,* and *S. pombe.*
- In WT nuclear DNA of *S. cerevisiae, S. paradoxus,* and *S. pombe,* there is no specific bias to rNMP that is most incorporated, whereas we observe rG to be the least incorporated. In RNase H2-null cells of *S. cerevisiae* and *S. paradoxus,* rC is more abundant in nuclear DNA, whereas rA is abundant in *S. pombe.* rU is the least incorporated in RNase H2-null cells of *S. cerevisiae, S. paradoxus,* and *S. pombe*.
- The pattern of rNMPs in mitochondrial DNA from WT and RNase H2-null of *S. cerevisiae* and *S. paradoxus* are similar. We observed rC as most abundantly incorporated rNMPs, whereas in *S. pombe* rG is most abundantly incorporated. In all three yeast species, we observe rU as the least incorporated rNMP in mitochondrial DNA.
- RNase H2 has a stronger activity on rNMPs in nuclear DNA as compared to mitochondrial DNA.

- rNMPs are incorporated within the G+C region of *S. cerevisiae* and *S. paradoxus*. rUMP is the least incorporated rNMPs in all mitochondrial DNA and *rnh201*Δ nuclear DNA.
- The low frequency of rUMPs is not due to the removal of rUMPs by UNG1 of the BER pathway.
- Yeast, mouse and human RNase H2 does not cleave abasic rNMPs, whereas human APE1 can recognize and cleave an abasic rNMP.
- Abasic rNMP cleavage is specific to APE1 as the use of compound #3 drastically reduced the cleavage activity of APE1.
- Identified a new role of APE1 in removing abasic rNMPs in genomic DNA.

6.2 Impact and future perspectives

There are abundant evidences that demonstrate the deleterious effects of misincorporation of rNMPs, but only a very few positive consequences of having rNMPs in DNA are known. Considering that rNMPs are the most frequent non-standard nucleotides, studying the rNMP profile of genomic DNA will help us better understanding both helpful and harmful effects of rNMPs embedded in DNA. Having developed an rNMP mapping technique that directly captures the rNMPs, ribose-seq, we set out to modify and optimize our current ribose-seq method to generate a powerful and effective system for mapping rNMPs in DNA of any kind. Using this modified method, we reveal the identity and distribution of rNMPs in genomic DNA of RNase H2 wild-type and several RNase H2 mutant cells of *S. cerevisiae*, *S. paradoxus*, and *S. pombe*. From our ribose-seq analysis, we noticed a difference in the overall content of rNMPs incorporated in nuclear and mitochondrial DNA

between WT and RNase H2-null cells. On normalizing the relative frequencies of nucleotides, we observed that rC and rG are more frequently incorporated in the genome (in most cases). rA and particularly rU, are the least incorporated rNMPs. This is especially interesting in mitochondria. Despite mitochondrial DNA being A+T rich, we found higher incorporation of rC and rG in G+C rich regions. This was observed in wild-type and RNase H2 mutants of budding yeasts, and not fission yeast. In *rnh201*Δ cells, we found a higher incorporation of rC, followed by rG, rA and low levels of rU incorporation. Additionally, using our analysis, we plan on identifying hotspots of rNMP incorporation in the genome.

In the absence of RER, Top1 has been shown to cleave rNMPs in DNA [35, 57]. rNMP-induced genome instability has been shown to result in 2-5 bp deletions that are dependent on Top1 activity. Using ribose-seq, we can identify where top1-dependent deletions occur in the genome after rNMP incorporation and persistence. It has been suggested that Top1-mediated cleavage at rNMPs may be related to a role in transcription rather than actually removing rNMPs from the genome. Therefore, comparing the profile of rNMPs in WT, RNase H2-null, Top1-null, and both RNase H2-null and Top1-nul cells can help identify sites that are target of Top1 versus RNase H2 in genomic DNA. As discussed in section 1.5.2 of the Introduction when Top1 cleaves at rNMPs that are part of a short DNA repeated sequence, it can lead to small deletions in DNA. Thus, it will be relevant to identify hotspot sites of rNMP incorporation, which could cause Top1-dependent mutations, not only in nuclear but also in mitochondrial DNA. Yeast has a Top1 that is equally active in nuclear and mitochondria, whereas there is a specialized Top1mt in mitochondria of human cells [181]. Very little is known about rNMP removal in mitochondria, hence it will be interesting to study if Top1 has a role in removing rNMPs in mitochondria.

In the course of my research, the modifications we made to ribose-seq have transformed the original protocol mainly applicable to RNase H2-defective cells into a robust technique. We improved the capacity of ribose-seq to capture rNMPs up to a factor of 1,000! Now, ribose-seq allows us to explore rNMP incorporation into DNA of potentially any cell type of any organisms under any conditions. For example, using ribose-seq we can identify rNMP incorporation in cells undergoing high level of DNA damage, such as oxidative stress. Moreover, it would be of interest to study rNMP incorporation in higher eukaryotic systems. As mutations in human RNase H2 subunits are associated with the neurodegenerative syndrome of AGS, ribose-seq mapping of system having partial loss-of-function RNase H2 can provide insights into how rNMP distribution affects AGS [54].

In addition to performing genome wide mapping, using our ribose-seq method, we can perform deep sequencing on a specific region to detect rNMPs. For example, we can construct ribose-seq libraries from telomeric, centromeric or any specific region of normal and diseased cell lines, and compare the rNMP profile in those regions between normal and diseased state. This would provide us information on regions that are more vulnerable to rNMP incorporation. Also, performing deep sequencing of ribose-seq libraries can help us characterize presence of rNMPs in repeated regions. At least 55% of the human genome comprise of repetitive elements, such as retrotransposons, simple sequence repeats and satellite DNA sequences [182]. As these elements play important roles in shaping genomes during evolution, and have been implicated in etiology of many human diseases, performing ribose-seq analysis to understand the identity and location of rNMPs in repeat sequences by deep-sequencing will help us gain information on the potential role of rNMPs in these repeated sequences.

It is well known that genomic instability is a hallmark of cancer; hence, analysis of rNMP spectra in cancer cells will reveal any correlation between rNMPs and cancer. On performing DNA-seq and ribose-seq on specific cancer cells/tissues, we can compare the DNA-seq profile to that of rNMP profile from ribose-seq to look for specific mutations in cancer cells/tissues. This could help in development of newer and more effective strategies for cancer treatment. Therefore, characterizing the rNMP profile in certain types of diseased cells can help us in better understanding the molecular mechanisms of these diseases, and possibly aid to develop effective tools for diagnostic and treatment of these diseases.

As of now, we are able to capture rNMPs from a heterogeneous population of cells. It will be interesting to determine the rNMP profile in a single cell. As eukaryotic cell populations are heterogeneous in nature, performing single-cell ribose-seq analysis will help us to understand mechanisms that are not seen when studying a bulk population of cells. This will prove to be very useful in mapping rNMPs within tumors. As tumors likely contain multiple sub-clones of cancer cells, we may be able to perform single-cell analysis of rNMPs. Using our analysis, we can accurately determine the pattern of rNMPs in a single-cell, as well as obtain genome wide reads of rNMPs.

Although using ribose-seq (and other methods) we can effectively capture rNMPs embedded in DNA, currently we cannot distinguish between a single rNMP or rNMPs coming from stretches of rNMPs in DNA. Due to alkali treatment, we can only capture the 5'-rNMP in a stretch of rNMPs. Therefore, having a RNase H2 mutant that cannot recognize stretches of rNMPs but can cleave at single rNMPs in DNA (reverse activity of RNase H2 RED mutant), and comparing the rNMP spectrum of these mutants both in the

presence and in the absence of RNase H1, which also targets stretches of rNMPs in DNA, we can identify the location of single rNMPs versus stretches of rNMPs.

Among the many rNMPs that are misincorporated in a given genome, not only canonical rNMPs are incorporated but also modified rNMPs such as abasic rNMPs might be found in DNA [27, 149]. rNMPs are susceptible to oxidative damage which could result in a significant generation of abasic rNMPs [8]. The role of RNase H2 in recognizing and cleaving rNMPs is well established but nothing was known about pathways that can cleave abasic rNMPs in DNA. In this research project, we demonstrated that an abasic rNMP embedded in DNA is targeted by APE1 of BER rather than RER in eukaryotic systems. We found that RNase H2 from yeast, mouse and human cell are unable to process abasic rNMP. We also showed that UNG1 of BER pathway did not target rUMP embedded in DNA. Our results highlight a new role of APE1 in repairing abasic rNMPs embedded in DNA. As APE1 overexpression is observed in different types of cancers [166, 167], such as ovarian, lung and breast cancers, it would be interesting to determine if there is any correlation between overexpression of APE1 and presence of abasic rNMPs. As we are understanding more and more about the role of rNMPs in genomic DNA, it would be of significance to study the consequence of abasic rNMPs, and other modified rNMPs in DNA.

We believe that our work will serve as a foundation for future studies aimed in better understanding the physiological and pathological significance of rNMPs in DNA. Our discovery of a novel function of BER pathway highlights the importance of exploring the capacity of additional DNA repair mechanisms to target rNMPs in DNA.

APPENDIX A. Supplementary Materials for Chapter 2

Table A.1. All strains used in this chapter

a

Strain	Relevant genotype	Source
KK-100	*MATα ade5-1 lys2-14A trp1-289 his7-2 leu2-3,112 ura3-52 rnh201*Δ::*hygMX4*	this study
KK-30	*ho*Δ *hml*Δ::*ADE1 MATa-inc hmr*Δ::*ADE1 ade1 leu2-3,112 lys5 trp1*::*hisG ura3-52 leu2*::*HOcs mata*Δ::*hisG rnh201*Δ::*hygMX4*	this study
KK-174	KK-100 *rnh1*Δ::*kanMX4*	this study
KK-125	KK-30 *rnh1*Δ::*kanMX4*	this study
KK-164	KK-125 *ung1*Δ::*natMX4*	this study
KK-170	KK-30 *pol2-M644G*	this study
KK-107	KK-100 *pol2-4*	this study
KK-120	KK-100 *pol3-5DV*	this study

b

Strain	Relevant genotype	Source
FRO-767,768	*ho*Δ *hml*Δ::*ADE1 MATa-inc hmr*Δ::*ADE1 ade1 leu2-3,112 lys5 trp1*::*hisG ura3-52 ade3*::*GAL*::*HO leu2*::*HOcs mata*Δ::*hisG*	Storici *et al*., 2007[43]
FRO-984,985	FRO-767,768 *rnh201*Δ::*kanMX4*	Storici *et al*., 2007[43]
KK-158,159	FRO-767,768 *ung1*Δ::*hygMX4*	this study

Yeast strains used in (**a**) ribose-seq library construction and (**b**) DSB repair assay with rNMP-containing oligos.

Table A.2. List of oligos used in this chapter

Name	Length (nt)	Sequence (5'-3') with end modifications	Purification
Lig.47.D	47	CCCGAGTGTGATCATCTGGTCGCTGGGGAATGAGTCAGGCCACGGCG	PAGE
Lig.47.R	47	CCCGAGTGTGATCATCTGGTCGCTGGGGAATrGAGTCAGGCCACGGCG	PAGE
Lig.30.rA	30	NNNNNNNNNNNNNNNNNNNNrANNNNNNNN	PAGE
Lig.30.rG	30	NNNNNNNNNNNNNNNNNNNNrGNNNNNNNN	PAGE
Lig.30.rU	30	NNNNNNNNNNNNNNNNNNNNrUNNNNNNNN	PAGE
Lig.30.rC	30	NNNNNNNNNNNNNNNNNNNNrCNNNNNNNN	PAGE
Adaptor.L	87	P-NNNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAG AGGGAGTTCAGACGTGTGCTCTTCCGATCTAGCCAGCGCAGACCGTGAG GT	PAGE
Adaptor.S	20	P-CCTCACGGTCTGCGCTGGCT-Am	Desalted
PCR.1.Index1	63	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGT GTGCTCTTCCGATC	Desalted
PCR.1.Index2	63	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGT GTGCTCTTCCGATC	Desalted
PCR.1.Index3	63	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGT GTGCTCTTCCGATC	Desalted
PCR.1.Index4	63	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGT GTGCTCTTCCGATC	Desalted
PCR.2	58	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCT TCCGATCT	Desalted
ByTemp.rC	46	NNNNNNNrCNNNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAG	PAGE
ByTemp.rU	46	NNNNNNNrUNNNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAG	PAGE
ByPrim	30	CTCTTTCCCTACACGACGCTCTTCCGATCT	PAGE
LEU2.D	60	TTAGGTGCTGTGGGTGGTCCTAAATGGGGATCCGGTAGTGTTAGGCCTG AACAAGGTTTA	Desalted
LEU2.rG	60	TTAGGTGCTGTGGGTGGTCCTAAATGGGGATCCGGTAGTrGTTAGGCCTG AACAAGGTTTA	Desalted
LEU2.dU	60	TTAGGTGCTGTGGGTGGTCCTAAATGGGGATCCGGTAGTGUTAGGCCTG AACAAGGTTTA	Desalted
LEU2.rU	60	TTAGGTGCTGTGGGTGGTCCTAAATGGGGATCCGGTAGTGrUTAGGCCTG AACAAGGTTTA	Desalted
LEU2.3	20	ATGTCTGCCCCTAAGAAGAT	Desalted
LEU2.6	20	TGCCAAAGAATAAGGTCAAC	Desalted

Name, length, and sequence of oligos used in this study are described. The purification type and the specific experiments in which the oligos were used are indicated. Ribonucleotides are in red, preceded by 'r'. End modifications of phosphate and amino groups are indicated by 'P' and 'Am', respectively. All PAGE-purified oligos were synthesized by Thermo Scientific Dharmacon with exceptions for Lig.47.D and Adaptor.L,

which were synthesized by Life Technologies and IDT, respectively. All desalted oligos were synthesized by Eurofins Genomics.

Table A.3 Results of 3' base bias for AtRNL ligation

a

Base	Circular	Dimer	Circular dimer
A	48% (44–49)	4.3% (1.8–7.1)	1.5% (0.71–2.1)
G	47% (44–48)	4.0% (2.6–5.8)	1.5% (0.88–2.3)
U	47% (45–49)	4.4% (2.1–5.1)	1.5% (0.73–2.0)
C	47% (44–49)	4.5% (1.9–5.0)	1.4% (0.59–1.8)

b

P value	G	U	C
A	0.4857	0.8857	1.0000
G	–	0.8857	0.6857
U	–	–	1.0000

c

Base	Circular	Dimer	Circular dimer
A	27% (23–31)	1.7% (1.5–2.3)	0.53% (0.45–0.74)
G	27% (24–28)	1.9% (1.5–2.5)	0.54% (0.39–1.1)
U	29% (24–30)	2.3% (1.3–2.7)	0.64% (0.48–1.2)
C	29% (25–32)	1.8% (1.7–2.0)	0.53% (0.47–0.80)

d

P value	G	U	C
A	1.0000	1.0000	0.6857
G	–	0.3429	0.2000
U	–	–	0.8857

(**a**) Levels of AtRNL ligation in reaction conditions described in Supplementary Figure 1 are expressed as median percentage and range (in parentheses) from four independent reactions. (**b**) Mann-Whitney *U*-test was performed for statistical analysis, and *P* values are displayed, all greater than 0.05. (**c**) AtRNL ligation was performed with reduced 200 nM AtRNL, instead of 1 μM, to compare the levels of ligation when the reactions were incomplete. Median percentages and ranges (in parentheses) from four independent

reactions are displayed. **d,** Mann-Whitney *U*-test was performed for statistical analysis, and *P* values are displayed, all greater than 0.05. No 3′ base bias was observed for AtRNL ligation.

Table A.4 Ribose-seq coverage for each library in this study

Ribose-seq library	Coverage (aligned reads/kb)	
	Nuclear	Mitochondrial
rnh201 (KK-100)	0.449	19.5
rnh201 (KK-100, EconoTaq)	0.883	47.8
rnh201 (KK-30)	0.149	8.42
rnh1 rnh201 (KK-174)	0.149	9.92
rnh1 rnh201 (KK-125)	0.239	13.2
rnh1 rnh201 ung1 (KK-164)	0.269	42.2
pol2-M644G rnh201 (KK-170)	0.254	7.89
pol2-4 rnh201 (KK-107)	0.528	34.2
pol3-5DV rnh201 (KK-120)	0.510	33.9

Coverage is expressed as aligned reads per kb and does not reflect the relative abundance of rNMPs among different strains.

Table A.5 Absolute nucleotide frequencies of rNMPs and 3' flanking nucleotide

	Base	Position 0		Position +1	
		Nuclear	Mitochondrial	Nuclear	Mitochondrial
rnh201 (KK-100)	A	15.4%	25.6%	45.0%	45.8%
	C	44.0%	36.8%	22.4%	15.3%
	G	28.1%	19.0%	16.5%	5.8%
	U/T	12.5%	18.7%	16.1%	33.1%
rnh201 (KK-100, EconoTaq)	A	23.2%	38.2%	43.3%	43.5%
	C	35.4%	25.6%	19.3%	10.0%
	G	22.7%	14.5%	13.5%	6.2%
	U/T	18.7%	21.7%	23.9%	40.3%

rnh201 (KK-30)	A	20.4%	35.7%	47.5%	47.8%
	C	39.2%	28.3%	19.6%	11.1%
	G	27.5%	24.1%	14.4%	7.2%
	U/T	12.8%	11.9%	18.5%	33.8%
rnh1 rnh201 (KK-174)	A	17.1%	33.6%	44.9%	46.5%
	C	40.2%	27.0%	22.0%	11.8%
	G	27.7%	23.7%	15.0%	7.4%
	U/T	15.1%	15.7%	18.2%	34.2%
rnh1 rnh201 (KK-125)	A	20.1%	35.4%	45.2%	44.3%
	C	36.8%	28.6%	19.4%	12.1%
	G	29.7%	20.9%	15.0%	5.9%
	U/T	13.4%	15.1%	20.4%	37.7%
rnh1 rnh201 ung1 (KK-164)	A	24.3%	35.8%	44.3%	47.1%
	C	35.3%	30.2%	19.4%	13.2%
	G	26.5%	22.7%	15.3%	6.5%
	U/T	14.0%	11.3%	21.0%	33.2%
pol2-M644G rnh201 (KK-170)	A	19.5%	38.9%	52.2%	47.2%
	C	40.3%	28.6%	18.2%	11.5%
	G	26.5%	21.5%	13.0%	6.6%
	U/T	13.7%	11.0%	16.6%	34.7%
pol2-4 rnh201 (KK-107)	A	14.9%	21.9%	42.5%	46.2%
	C	40.2%	43.1%	22.0%	16.3%
	G	23.6%	16.3%	16.4%	6.4%
	U/T	21.2%	18.6%	19.2%	31.1%
pol3-5DV rnh201 (KK-120)	A	20.4%	30.0%	44.3%	45.3%
	C	37.1%	33.1%	19.4%	14.6%
	G	25.0%	16.7%	15.3%	6.1%
	U/T	17.5%	20.3%	21.0%	34.0%

Absolute nucleotide frequencies of nuclear and mitochondrial rNMPs and the nucleotide immediately downstream (position +1) from each ribose-seq library.

Table A.6 Results of rNMP bypass by Phusion polymerase

a

Base	Bypass probability
C	93% (93–93)
U	93% (92–94)

b

P value	U
C	0.6857

(**a**) Bypass probabilities reaction conditions described in Supplementary Figure 4 are expressed as median percentage and range (in parentheses) from four independent reactions. (**b**) Mann-Whitney *U*-test was performed for statistical analysis, and *P* value is displayed.

Table A.7 Results of DSB repair assay with rNMP-containing oligos

a

Oligo	WT	*rnh201*	*ung1*
LEU2.D	65% (55–75)	63% (55–65)	65% (55–70)
LEU2.rG	30% (20–40)	90% (75–100)	N/A
LEU2.dU	5.0% (0–10)	N/A	55% (45–60)
LEU2.rU	33% (25–45)	55% (45–65)	33% (30–40)

b

Oligo	*rnh201*	*ung1*
LEU2.D	0.5357	1.0000
LEU2.rG	0.0286	N/A
LEU2.dU	N/A	0.0294
LEU2.rU	0.0421	1.0000

(**a**) Data shown in **Figure 2.4** are presented here as median percentages of StuI-cut Leu^+ transformants from four independent transformations and ranges in parentheses. For each transformation, 20 Leu^+ transformants were selected for analysis. **b,** Mann-Whitney *U*-test was implemented for statistical analysis against the WT, and *P* values are displayed. N/A, not applicable because data are not available for comparison.

Table A.8 List of hotspots of rNMP incorporation within *S. cerevisiae* mitochondrial DNA, rDNA repeat, and *Ty1*

Position	Strand	Gene	Base	Number of rNMP reads								
				rnh201 (KK-100)	*rnh201* (KK-100, EconoTaq)	*rnh201* (KK-30)	*rnh1 rnh201* (KK-174)	*rnh1 rnh201* (KK-125)	*rnh1 rnh201 ung1* (KK-164)	*pol2-M644G rnh201* (KK-170)	*pol2-4 rnh201* (KK-107)	*pol3-5DV rnh201* (KK-120)
Chr M 39,224	W	*COB*	A	**30**	**136**	**17**	**24**	**34**	2	20	10	22
Chr XII[a] 453,839	C	*RDN25*	G	**97**	**45**	**18**	**30**	**22**	1	25	12	10
Chr IV[b] 650,383	C[b]	*Ty1*	A	**15**	**42**	**10**	**11**	**19**	46	20	19	110
Chr M 14,688	W	*COX1*	A	**7**	**49**	**5**	**10**	**14**	33	12	8	38
Chr M 14,739	W	*COX1*	A	**8**	**46**	**7**	**14**	**12**	0	5	0	13
Chr M 19,157	W	*COX1*	A	**15**	**73**	**4**	**5**	**22**	28	5	7	27

Hotspots of of rNMP incorporation were determined by finding positions of rNMPs within the locus of interest with ribose-seq signal greater than the mean plus three standard deviations for each library from *rnh201*Δ (KK-100), *rnh201*Δ (KK-100, EconoTaq), *rnh201*Δ (KK-30), *rnh1*Δ *rnh201*Δ (KK-174), and *rnh1*Δ *rnh201*Δ (KK-125) cells (in bold). Ribose-seq signal counts found in all other libraries are also shown.

[a]There are two rDNA repeats on Chr XII in the reference genome (sacCer2). Only the first repeat unit is shown as an example.

[b]Because of the presence of multiple copies of *Ty1* in the genome, *YDRCTy1-1* on Chr IV is shown as an example.

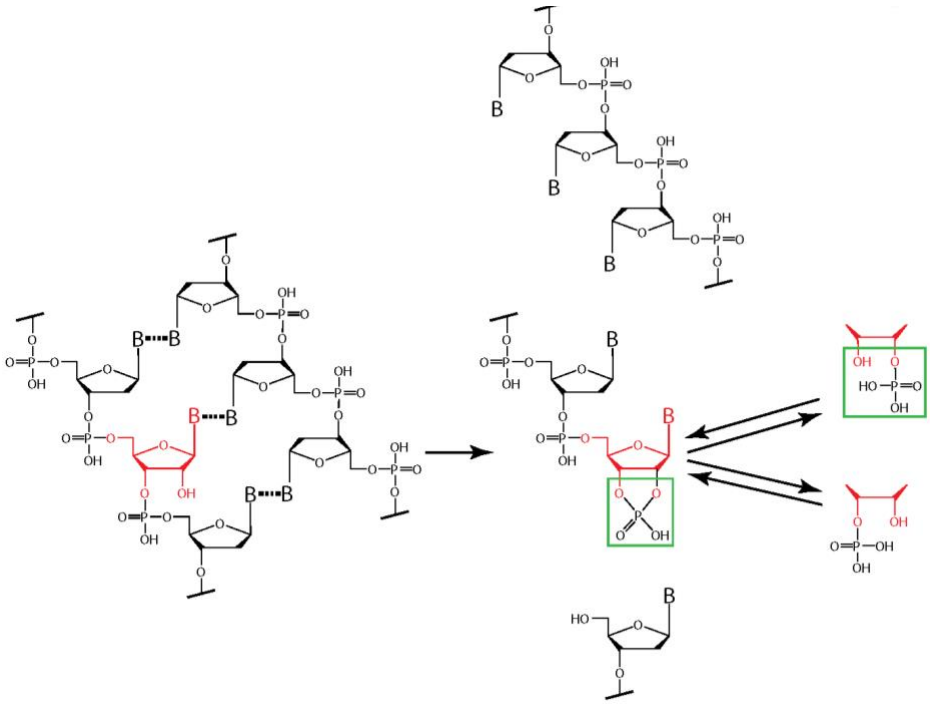


Figure A.1 Mechanism of alkaline cleavage of ribonucleotides in DNA. The ribonucleoside embedded in double-stranded DNA is in red. During alkaline treatment, DNA strands are denatured, and cleavage occurs at the rNMP site, generating a 2′,3′-cyclic phosphate end and an opposite 5′-hydroxyl end. The 2′,3′-cyclic phosphate is in equilibrium with 2′-phosphate and 3′-phosphate forms. Boxes in black indicate the 2′,3′-cyclic phosphate and 2′-phosphate DNA termini, which are substrates of AtRNL.

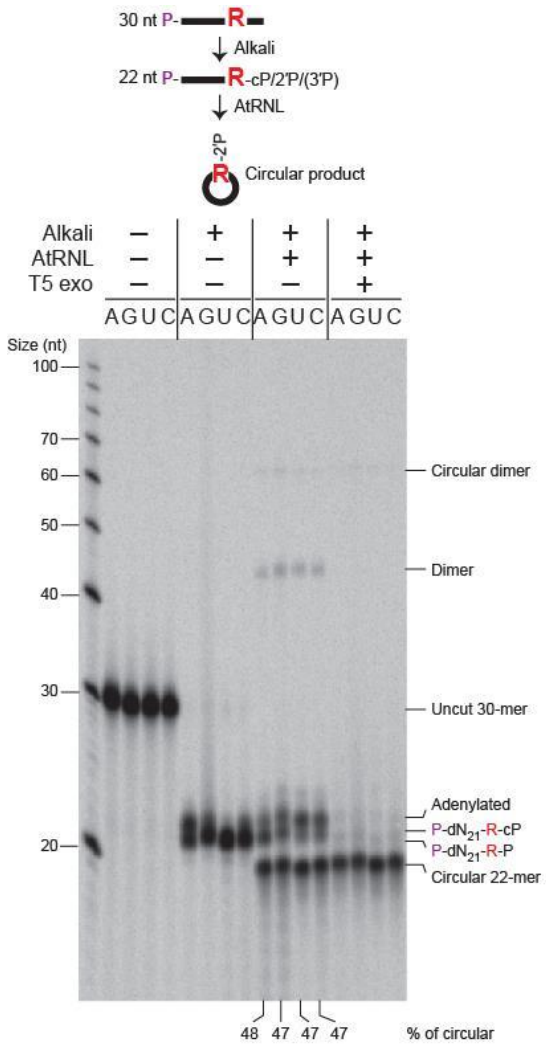


Figure A.2 3' bias for AtRNL ligation. Hot 5′-radiolabeled 30-nt DNA oligo with a single rNMP (either A, G, U, or C) in the 22nd position was mixed with cold equimolar 30-nt DNA oligos with rNMPs of 3 other bases in the 22nd positions. 5′-radiolabel is indicated by 'P' in purple. The mixture was treated with 0.3M NaOH for 2 hr at 55 °C and neutralized. 100 nM of alkali-cleaved products (25 nM of each base) were then incubated with 1 μM AtRNL in appropriate buffer (see Methods) for 1 hr at 30 °C. The resulting products were treated with T5 exonuclease for 2 hr at 37 °C. Aliquots were withdrawn after appropriate steps and quenched. The products were analyzed by urea-PAGE. The circular 22-mer migrates faster than the unligated, linear 22-mer. Only circular products were resistant to T5 exonuclease while all linear substrates/products were degraded. Median percentages of circular 22-mer formation from four independent reactions are displayed. See Supplementary Table 1 for more statistics. First left lane, ss DNA ladder. No 3′ base bias was observed for AtRNL ligation (see **Table A.3**). Self-ligation was preferred to dimerization with a shorter 22-nt substrate; however, with the shorter substrate, lower levels of linear dimers, which are not resistant to T5 exonuclease, and circular dimers were

observed. Increasing the length of the ss DNA substrate from 22 nt to 32 nt eliminated dimerization (**Figure 1.1a**).

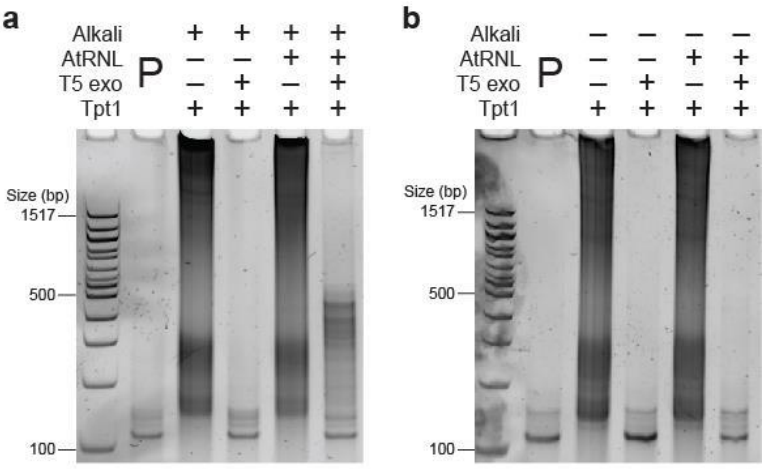


Figure A.3 Ribose-seq library from genomic DNA of *S. cerevisiae rnh201Δ* (KK-100) cells. Appropriate PCR products were analyzed by PAGE. 'P' indicates primers-only. No amplification product was observed when either (a) AtRNL ligation step or (b) alkali treatment was omitted. Tpt1 denotes the step of 2'-phosphate removal at the ligation junction in **Figure 1.1a**. First left lane, ds DNA ladder.

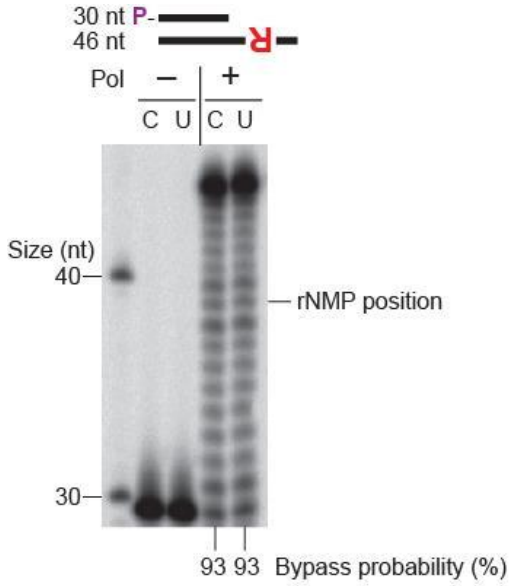


Figure A.4 Bypass of a single rNMP by Phusion DNA polymerase. 5′-radiolabeled 30-nt primer, ByPrim (**Table A.2**), was annealed to the 46-nt template oligo containing either rCMP (ByTemp.rC) or rUMP (ByTemp.rU) in the 8th position. 100 nM of annealed substrate was incubated with 0.2 units of Phusion High-Fidelity DNA Polymerase (NEB) and 2 mM dNTPs in appropriate buffer (see Methods) for 30 sec at 72 °C. The reactions were quenched and analyzed by urea-PAGE. Median bypass probabilities from four independent reactions are shown. See Supplementary Table 5 for more statistics. First left lane, ss DNA ladder. The primer extension assay showed no significant difference between bypass efficiency over rUMP and rCMP by Phusion DNA polymerase (**Table A.6**).

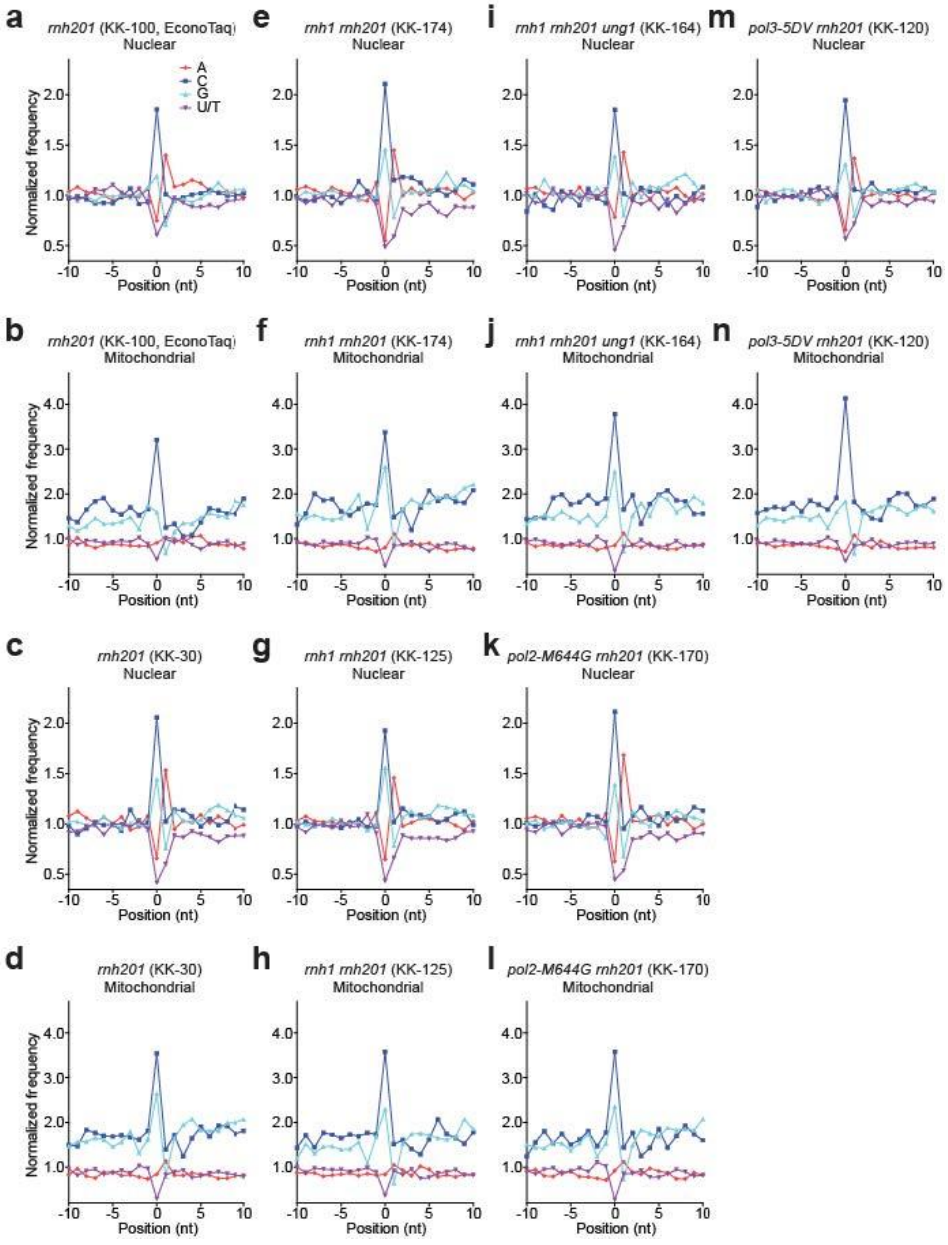


Figure A.5 Normalized frequency of nucleotides surrounding the rNMP sites. Normalized frequency of nucleotides relative to (**a**) nuclear and (**b**) mitochondrial mapped positions of sequences from ribose-seq library, PCR-amplified with EconoTaq DNA Polymerase (Lucigen), of genomic DNA from *S. cerevisiae rnh201*Δ (KK-100) cells. Position 0 corresponds to the rNMP. Negative and positive numbers (from -10 to -1 and 1 to 10) correspond to upstream and downstream positions from the rNMP, respectively. Frequencies were normalized to either nuclear or mitochondrial genomic mononucleotide frequencies. Normalized frequency of nucleotides relative to (**c**) nuclear and (**d**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae rnh201*Δ (KK-30) cells. Normalized frequency of nucleotides relative to (**e**) nuclear and (**f**) mitochondrial mapped positions of sequences from ribose-seq library

of genomic DNA from *S. cerevisiae rnh1*Δ *rnh201*Δ (KK-174) cells. Normalized frequency of nucleotides relative to (**g**) nuclear and (**h**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae rnh1*Δ *rnh201*Δ (KK-125) cells. Normalized frequency of nucleotides relative to (**i**) nuclear and (**j**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae rnh1*Δ *rnh201*Δ *ung1*Δ (KK-164) cells. Normalized frequency of nucleotides relative to (**k**) nuclear and (**l**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae pol2-M644G rnh201*Δ (KK-170) cells. Normalized frequency of nucleotides relative to (m) nuclear and (n) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae pol3-5DV rnh201*Δ (KK-120) cells.

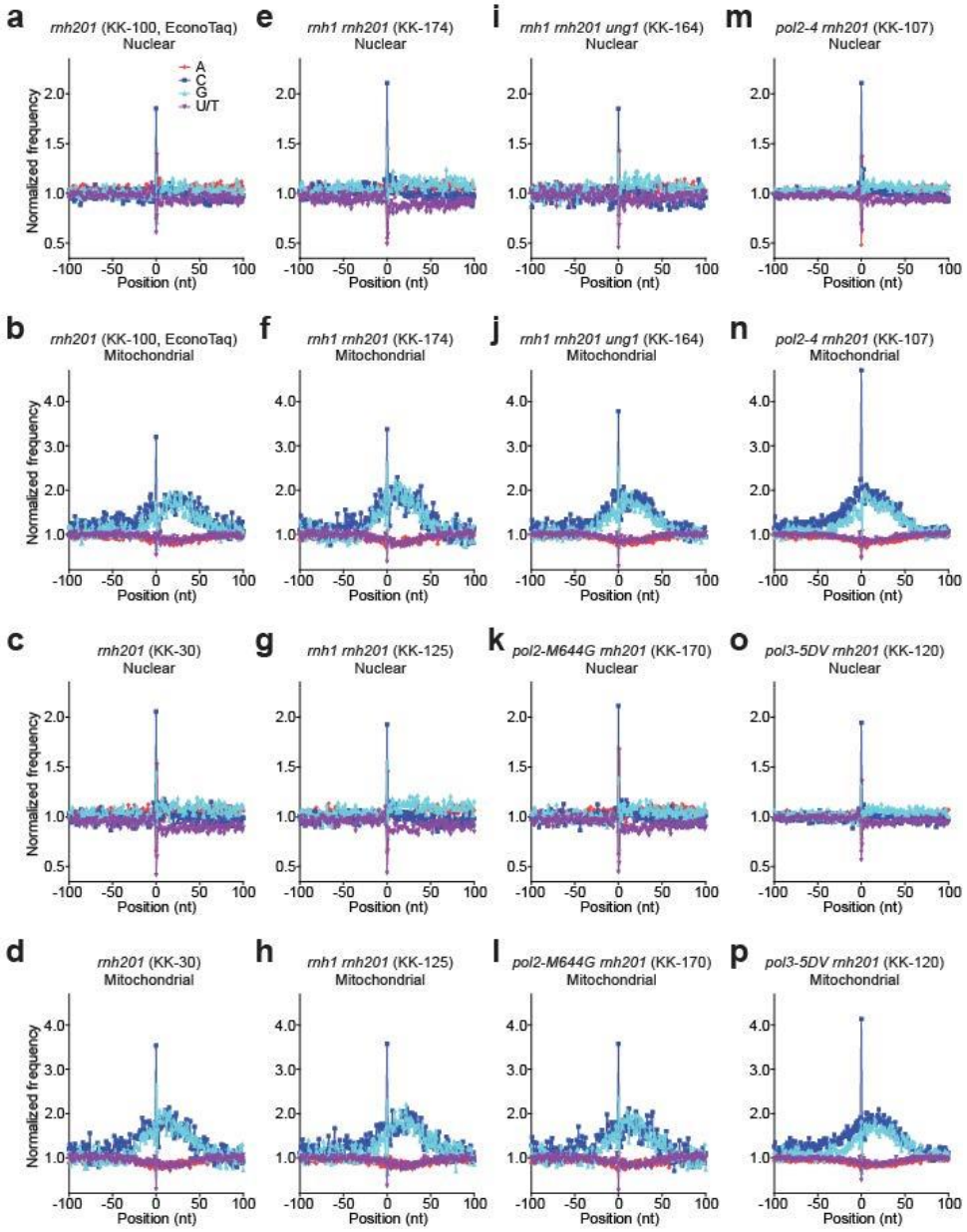


Figure A.6 Zoom-out of normalized frequency of nucleotides surrounding the rNMP sites. Normalized frequency of nucleotides relative to (**a**) nuclear and (**b**) mitochondrial mapped positions of sequences from ribose-seq library, PCR-amplified with EconoTaq DNA Polymerase (Lucigen), of genomic DNA from *S. cerevisiae rnh201*Δ (KK-100) cells. Position 0 corresponds to the rNMP. Negative and positive numbers (from -100 to -1 and 1 to 100) correspond to upstream and downstream positions from the rNMP, respectively. Frequencies were normalized to either nuclear or mitochondrial genomic mononucleotide frequencies. Normalized frequency of nucleotides relative to (**c**) nuclear and (**d**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae rnh201*Δ (KK-30) cells. Normalized frequency of nucleotides relative to (**e**) nuclear and (**f**) mitochondrial mapped positions of sequences from ribose-seq library

of genomic DNA from *S. cerevisiae rnh1*Δ *rnh201*Δ (KK-174) cells. Normalized frequency of nucleotides relative to (**g**) nuclear and (**h**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae rnh1*Δ *rnh201*Δ (KK-125) cells. Normalized frequency of nucleotides relative to (**i**) nuclear and (**j**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae rnh1*Δ *rnh201*Δ *ung1*Δ (KK-164) cells. Normalized frequency of nucleotides relative to (**k**) nuclear and (**l**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae pol2-M644G rnh201*Δ (KK-170) cells. Normalized frequency of nucleotides relative to (**m**) nuclear and (**n**) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae pol2-4 rnh201*Δ (KK-107) cells. Normalized frequency of nucleotides relative to (**o**) nuclear and (p) mitochondrial mapped positions of sequences from ribose-seq library of genomic DNA from *S. cerevisiae pol3-5DV rnh201*Δ (KK-120) cells.

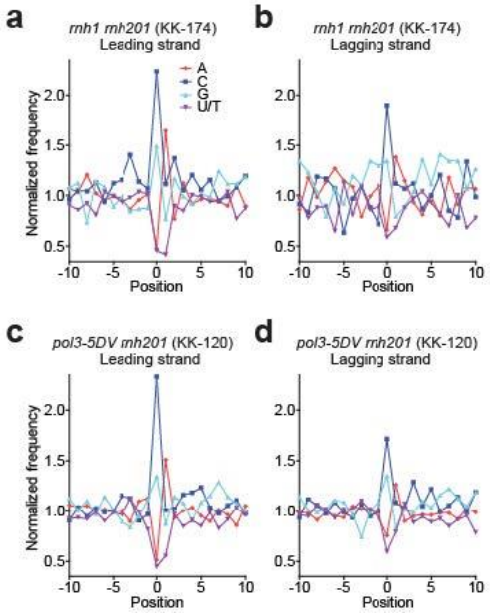


Figure A.7 Normalized frequency of nucleotides surrounding the rNMP sites on leading and lagging strands. Normalized frequency of nucleotides relative to mapped positions of sequences in (**a**) leading and (**b**) lagging strands from ribose-seq library of genomic DNA from *S. cerevisiae rnh1*Δ *rnh201*Δ (KK-174) cells. Position 0 corresponds to the rNMP. Negative and positive numbers (from -10 to -1 and 1 to 10) correspond to upstream and downstream positions from the rNMP, respectively. ARSs with T_{rep} of no longer than 25 min were selected with flanking size of 10 kb. Frequencies were normalized to genomic mononucleotide frequencies of either leading or lagging strand of the selected ARSs and flanking size. Normalized frequency of nucleotides relative to mapped positions of sequences in (**c**) leading and (**d**) lagging strands from ribose-seq library of genomic DNA from *S. cerevisiae pol3-5DV rnh201*Δ (KK-120) cells.

APPENDIX B. Supplementary Information for Chapter 3

Table B.1 List of single strand (ss) oligonucleotides used in this study

Name	Length	5' label and sequence (5'-3')
ss_dG_40	40	**^{32}P**-AGTGTGGATGCCGTGAAGA**dG**ATCATCAAGGTTATCGCCTC
ss_rG_40	40	**^{32}P**-AGTGTGGATGCCGTGAAGA**rG**ATCATCAAGGTTATCGCCTC
ss_dF_40	40	**^{32}P**-AGTGTGGATGCCGTGAAGA**dF**ATCATCAAGGTTATCGCCTC
ss_rF_40	40	**^{32}P**-AGTGTGGATGCCGTGAAGA**rF**ATCATCAAGGTTATCGCCTC
ss_dG	25	**IRDye700**-GGATCCGGTAGT**dG**TTAGGCCTGAAC
ss_rG	25	**IRDye700**-GGATCCGGTAGT**rG**TTAGGCCTGAAC
ss_rOH	25	**Cy5**-GGATCCGGTAGT**rOH**TTAGGCCTGAAC
ss_d8oxoG	25	**IRDye700**-GGATCCGGTAGT**d8oxoG**TTAGGCCTGAAC
ss_r8oxoG	25	**IRDye700**-GGATCCGGTAGT**r8oxoG**TTAGGCCTGAAC
ss_dF	26	**IRDye800**-AATTCACCGGTACC**dF**TCTAGAATTCG
ss_dC_40	40	GAGGCGATAACCTTGATGAT**C**TCTTCACGGCATCCACACT
ss_dC	25	GTTCAGGCCTAA**C**ACTACCGGATCC
ss_dA	25	GTTCAGGCCTAAA**A**CTACCGGATCC
ss_dC	26	CGAATTCTAGA**C**GGTACCGGTGAATT

Name, length, 5′ labeling and sequence of all single strand oligonucleotides used in this study are presented. All ss oligonucleotides were PAGE purified. Ribonucleotides are in red and bolded, and the corresponding deoxyribonucleotides are in blue and bolded. ^{32}P, IRDye700 phosphoramidite, IRDye800 phosphoramidite and Cyanine 5 dyes are indicated at the 5' end of each corresponding oligonucleotide. In the part colored in grey all single strand complementary oligonucleotides are listed with the nucleotide opposite to the ribonucleotide in bolded font.

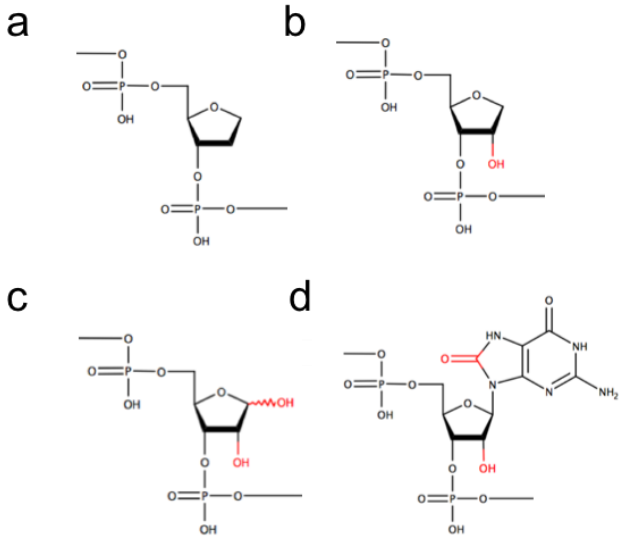


Figure B.1 Structure of modified bases incorporated in DNA oligomer. Schematic representation of the (**a**) deoxy- and (**b**) ribo- tetrahydrofuran, (**c**) ribose 1'OH abasic site and (**d**) ribose 8-Oxo-7,8-dihydro-guanosine- containing oligodeoxyribonucleotide.

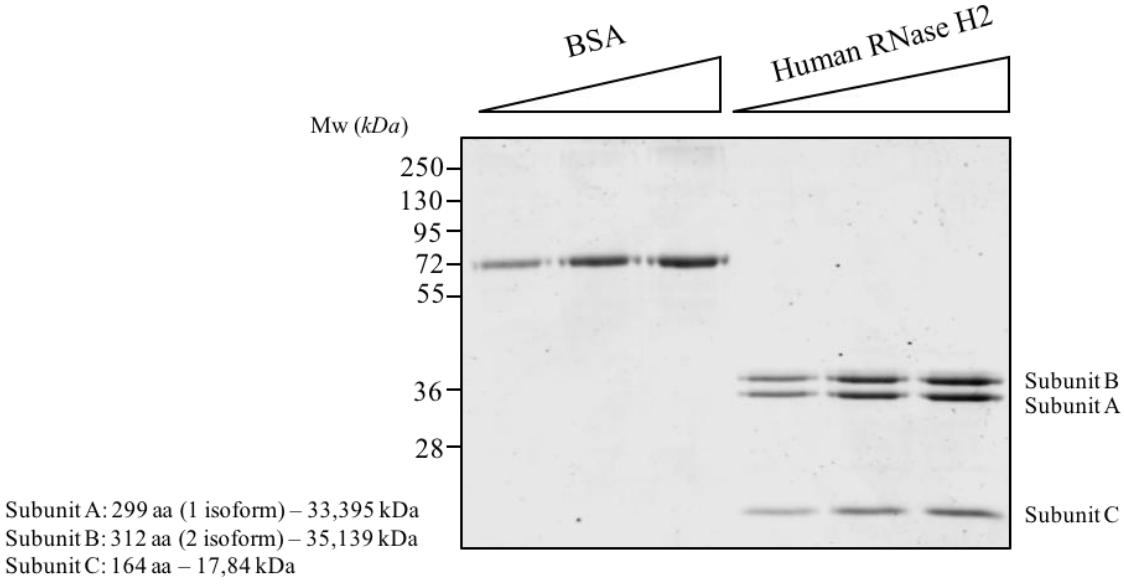


Figure B.2 Gel-quantification of RNase H2 recombination protein used in this study. Human recombinant RNase H2 protein was separated onto 10% SDS–PAGE gel followed by Coomassie staining. Bands corresponding to each protein were quantified and normalized on a standardization curve of Bovine Serum Albumin (BSA) protein. The molecular weight (Mw) expressed in kilodaltons (*kDa*) is shown on the right of each panel.

Subunit A, subunit B and subunit C corresponds to 299 aa (1 isoform, 33,395 kDa), 213 aa (2 isoform, 35,139 kDa), and 164 aa (1,784 kDa).

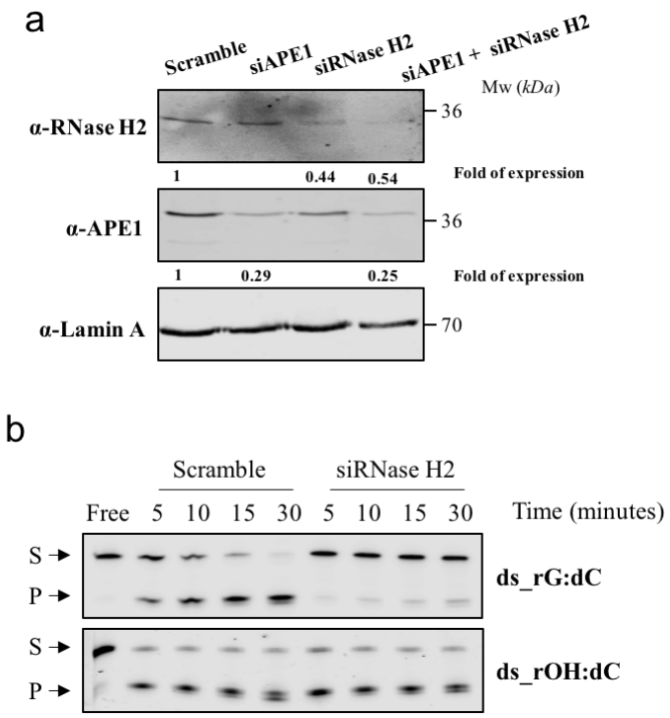


Figure B.3 Western blot analysis on nuclear HeLa cell extracts and its enzymatic activity on different substrates. (**a**) HeLa cells were transfected with specific siRNAs directed versus RNase H2 and APE1 proteins and compared to the Scramble control. Nuclear cell extracts were separated onto 12% SDS–PAGE, and Western blot analysis was performed by using an anti-APE1 antibody and an anti-RNase H2 antibody. Lamin A was used as loading control. Expression levels for each condition was normalized to the Scramble and indicated under each corresponding lane. The molecular weight (Mw) expressed in kilodaltons (*kDa*) is shown on the right of each panel. (**b**) Representative denaturing polyacrylamide gel of oligonucleotides (0.25 µM) incision by nuclear HeLa cell extracts (NCE). In order to discriminate the activity of RNase H2, 500 ng of NCE, in which RNase H2 expression was previously knocked down through specific siRNA (indicated as siRNase H2), were tested in comparison to control cells (Scramble) at different time points, expressed in minutes. rG- containing oligonucleotide was used as positive control. Time points, expressed in minutes, are shown on the top of the figure. Enzymatic reaction was performed at 37°C in RNase H2-buffer. S indicates the substrate position, while P indicates the product position.

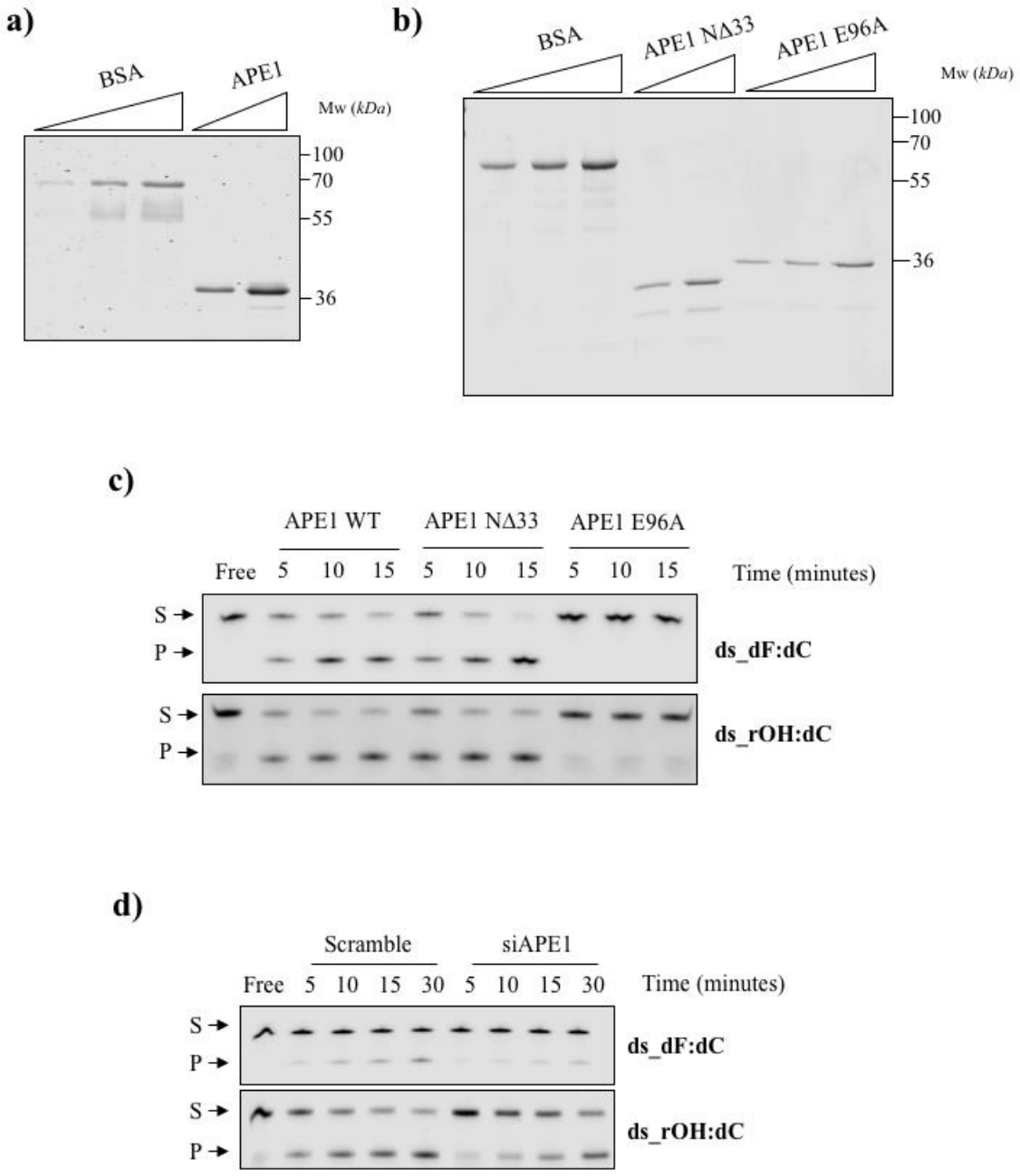
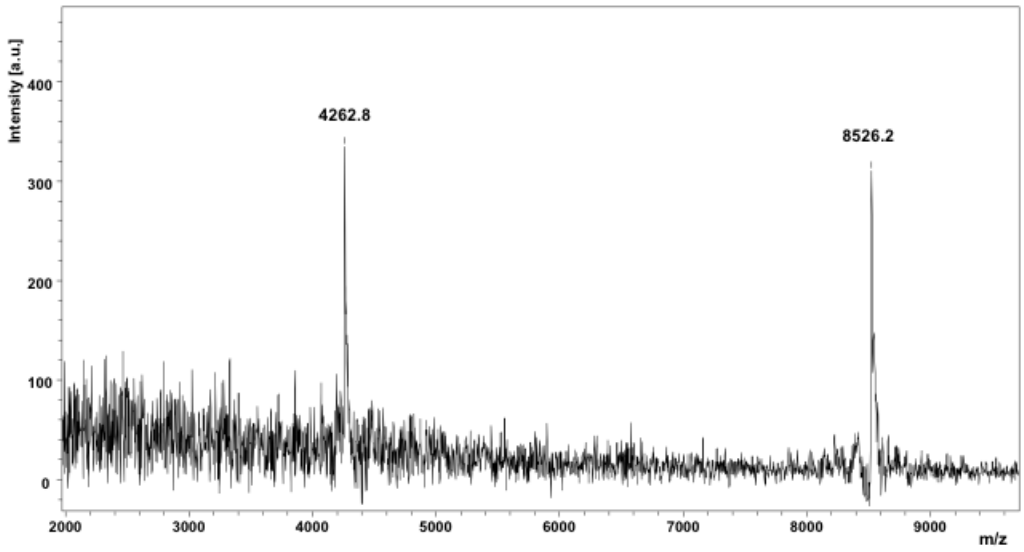


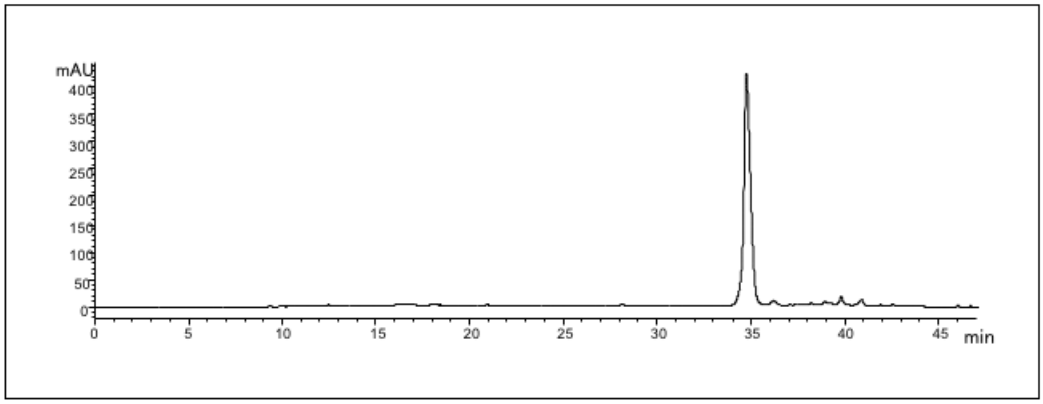
Figure B.4 Gel-quantification of APE1 WT and mutant recombinant proteins used in this study. Human recombinant APE1 WT (**a**) and mutants APE1 NΔ33 and APE1 E96A proteins (**b**) were separated onto 10% SDS–PAGE gel followed by Coomassie staining. Bands corresponding to each protein were quantified and normalized on a standardization curve of Bovine Serum Albumin (BSA) protein. The molecular weight (Mw) expressed in kilodaltons (*kDa*) is shown on the right of each panel. (**c**) Representative denaturing polyacrylamide gel of oligonucleotides (2.5 pmol) incision by different APE1 mutants NΔ33 and E96A (0.288 nM), in comparison to wild type APE1 (WT). The reaction was performed in APE1-buffer for different time points, expressed in minutes and shown on

the top of the figure, at 37°C. ds_dF:dC oligonucleotide was used as positive control. S indicates the substrate position while P indicates the product position. (**d**) Representative denaturing polyacrylamide gel of oligonucleotides (0.25 µM) incision by nuclear HeLa cell extracts (NCE). In order to discriminate the activity of APE1, 10 ng of NCE, in which APE1 expression was previously knocked down through specific siRNA (indicated as siAPE1), were tested in comparison to control cells (Scramble) at different time points, expressed in minutes and shown on the top of the figure. dF- containing oligonucleotide was used as positive control. Enzymatic reaction was performed at 37°C in APE1-buffer. S indicates the substrate position, while P indicates the product position.

a)


Intensity [a.u.]
400
300
200
100
0
4262.8
8526.2
2000
3000
4000
5000
6000
7000
8000
9000
m/z

b)


mAU
400
350
300
250
200
150
100
50
0
0
5
10
15
20
25
30
35
40
45
min

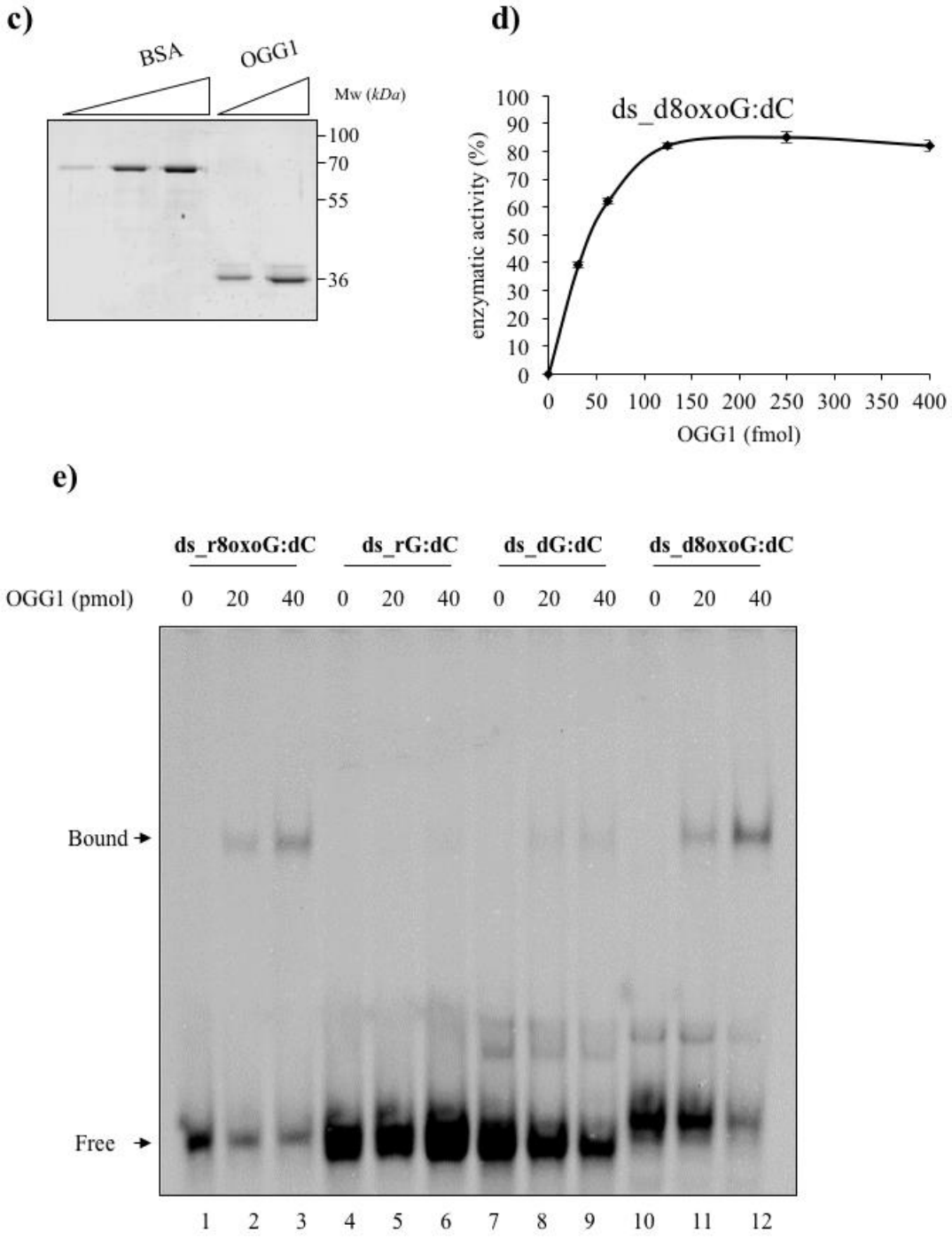


Figure B.5 Human OGG1 binds r8oxoG- containing oligonucleotide. (**a**) MALDI-TOF mass spectrum of the modified 25-mer oligodeoxyribonucleotide that contains a r8oxoG residue (calculated mass: 8528.9; measured mass: 8527.2). Here the classical spectrum for a pure oligonucleotide showing an unique fragment with two states of charge: Peak [M-H]- : 8526.2 and Peak [M-2H]2- : 4262.8 is reported for the oligonucleotide synthesized and used in all the experiments. (**b**) Reversed-phase chromatography of the labeled

modified oligonucleotide (Gradient of 0 to 35% of CH3CN in TEAA 10mM in 45min – detection 254nm). (**c**) Human recombinant OGG1 protein was separated onto 10% SDS–PAGE gel followed by Coomassie staining. Bands corresponding to each protein were quantified and normalized on a standardization curve of Bovine Serum Albumin (BSA) protein. The molecular weight (Mw) expressed in kilodaltons (*kDa*) is shown on the right of each panel. (**d**) Graph depicting dose response of OGG1 lyase activity on ds_d8oxoG:dC positive control oligonucleotide is shown. Data are expressed as mean ± SD of three independent technical replicas. (**e**) Representative native EMSA polyacrylamide gel of OGG1 binding on ds_r8oxoG:dC oligonucleotide (25 nM) is shown. ds_dG:dC and ds_rG:dC are used as negative controls whereas ds_d8oxoG:dC as used as positive control. The 'Bound' arrow denotes the retarded complex between OGG1 and the probe whereas the 'Free' arrow denotes the unbound substrate. Amounts of OGG1 protein, expressed in *pico* moles, are shown on the top of the figure. Reactions were performed as explained in "Material and Methods" section.

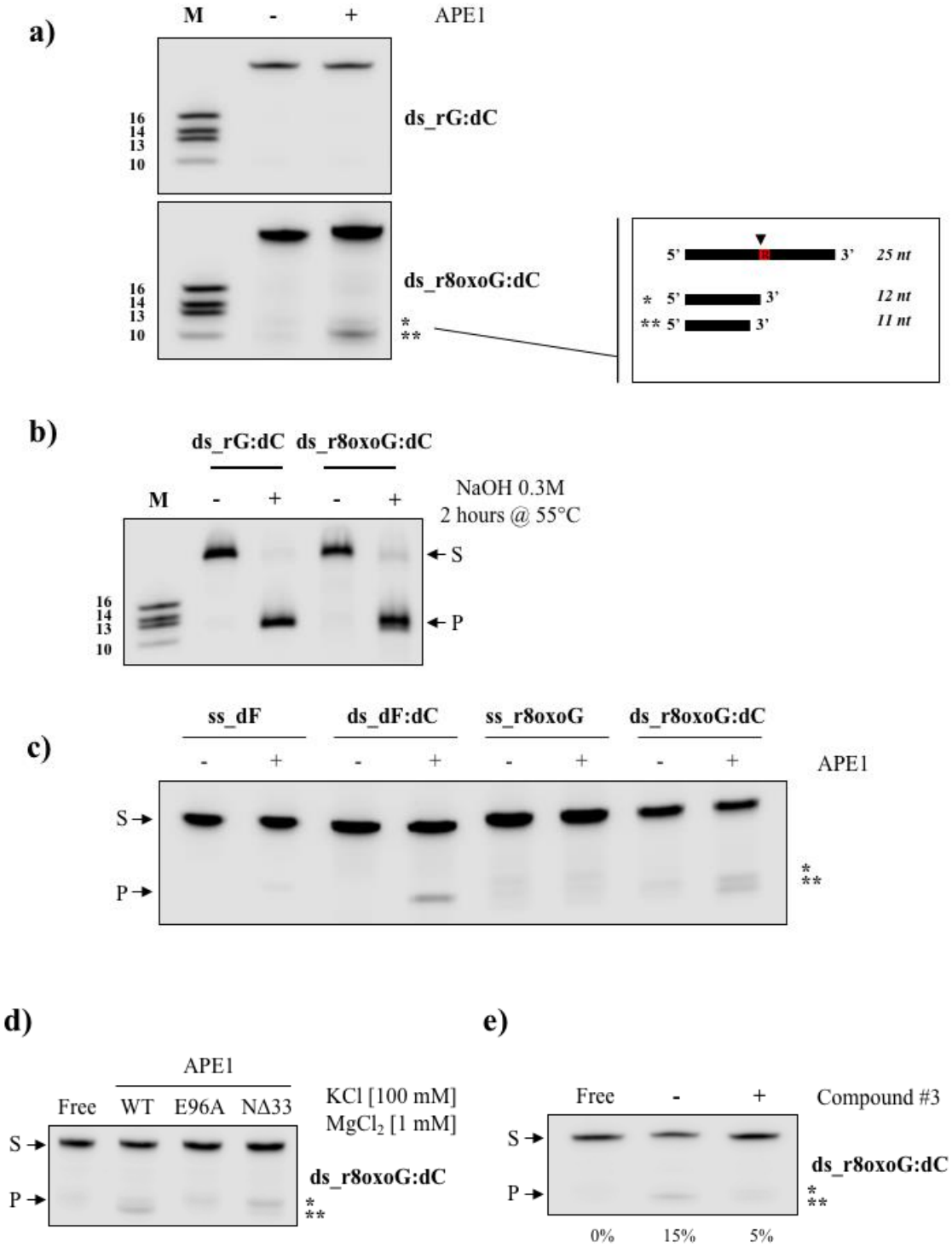
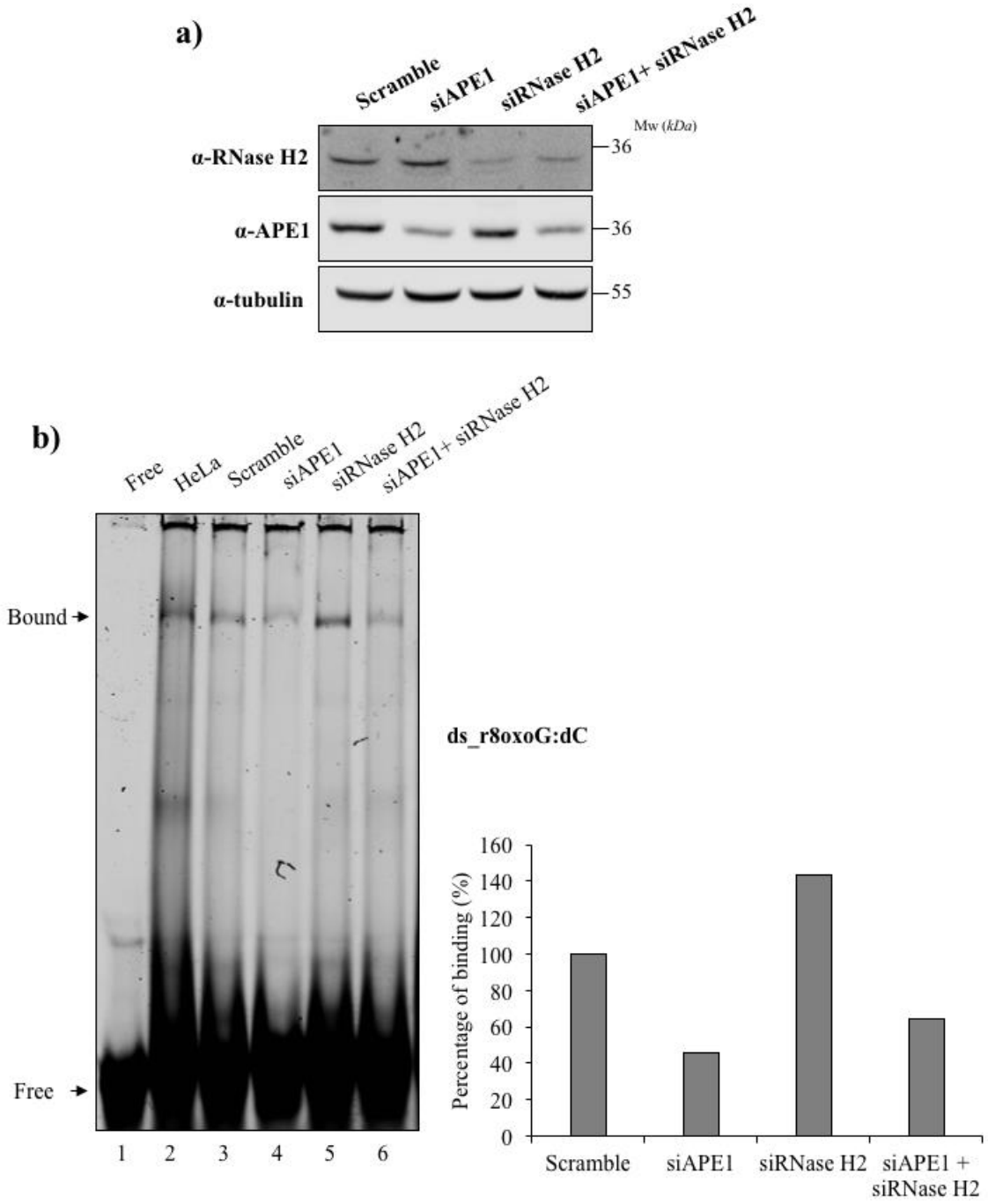


Figure B.6 Evaluation of the apparent molecular weight of the cleavage products by APE1. (**a**) Representative denaturing polyacrylamide gel of ds_rG:dC and ds_r8oxoG:dC

oligonucleotides (25 nM) incision by APE1 (50 fmol) in APE1 buffer for 30 minutes at 37°C. At the right of the bottom panel, a longer product of about 12 nucleotides is indicated by an asterisk whereas a smaller one of 11 nucleotides is indicated by a double asterisk. On the right side, a schematic representation of the cleavage products, showing the position of the ribonucleotide (red box with R) embedded in the DNA oligonucleotide and the APE1 cleavage on it, producing a longer product of about 12 nucleotides and a smaller one of 11 nucleotides. (**b**) Representative denaturing polyacrylamide gel of alkaline hydrolysis on ds_rG:dC and ds_r8oxoG:dC oligonucleotides (25 nM) upon treatment with 0.3 M NaOH for 2 hours at 55°C. On the left of the panel, marker shows the size of the products. (**c**) APE1 enzymatic activity on single strand (*ss*) versus double strand (*ds*) dF- and r8oxoG-containing oligonucleotides. 2 fmol and 50 fmol of APE1 protein were incubated with dF-containing oligonucleotide and r8oxoG-containing oligonucleotides for 10 and 30 minutes, respectively (*top*). APE1 enzymatic activity on single strand (*ss*) versus double strand (*ds*) dF- and rOH-containing oligonucleotides. 2 fmol of APE1 protein were incubated with oligonucleotides for 10 minutes (*bottom*). Reactions were prepared in APE1 buffer in a 10 µl of final volume at 37°C. S indicates the substrate position while P indicates the product position. (**d**) Representative denaturing polyacrylamide gel of APE1 mutants (5 nM), including E96A and NΔ33, on the ds_r8oxoG:dC oligonucleotide (25 nM). S indicates the substrate position, while P indicates the product position. Moreover, at the right of the panel, a longer product of about 12 nucleotides is indicated by an asterisk whereas a smaller one of 11 nucleotides is indicated by a double asterisk. (**e**) Representative denaturing polyacrylamide gel of APE1 (5 nM) incision in the presence of Compound #3 inhibitor. S indicates the substrate position, while P indicates the product position. Moreover, at the right of the panel, a longer product of about 12 nucleotides is indicated by an asterisk whereas a smaller one of 11 nucleotides is indicated by a double asterisk.


a)
Scramble
siAPE1
siRNase H2
siAPE1+ siRNase H2
Mw (kDa)
α-RNase H2
36
α-APE1
36
α-tubulin
55
b)
Free
HeLa
Scramble
siAPE1
siRNase H2
siAPE1+ siRNase H2
Bound
Free
1 2 3 4 5 6
ds_r8oxoG:dC
Percentage of binding (%)
160
140
120
100
80
60
40
20
0
Scramble
siAPE1
siRNase H2
siAPE1 + siRNase H2

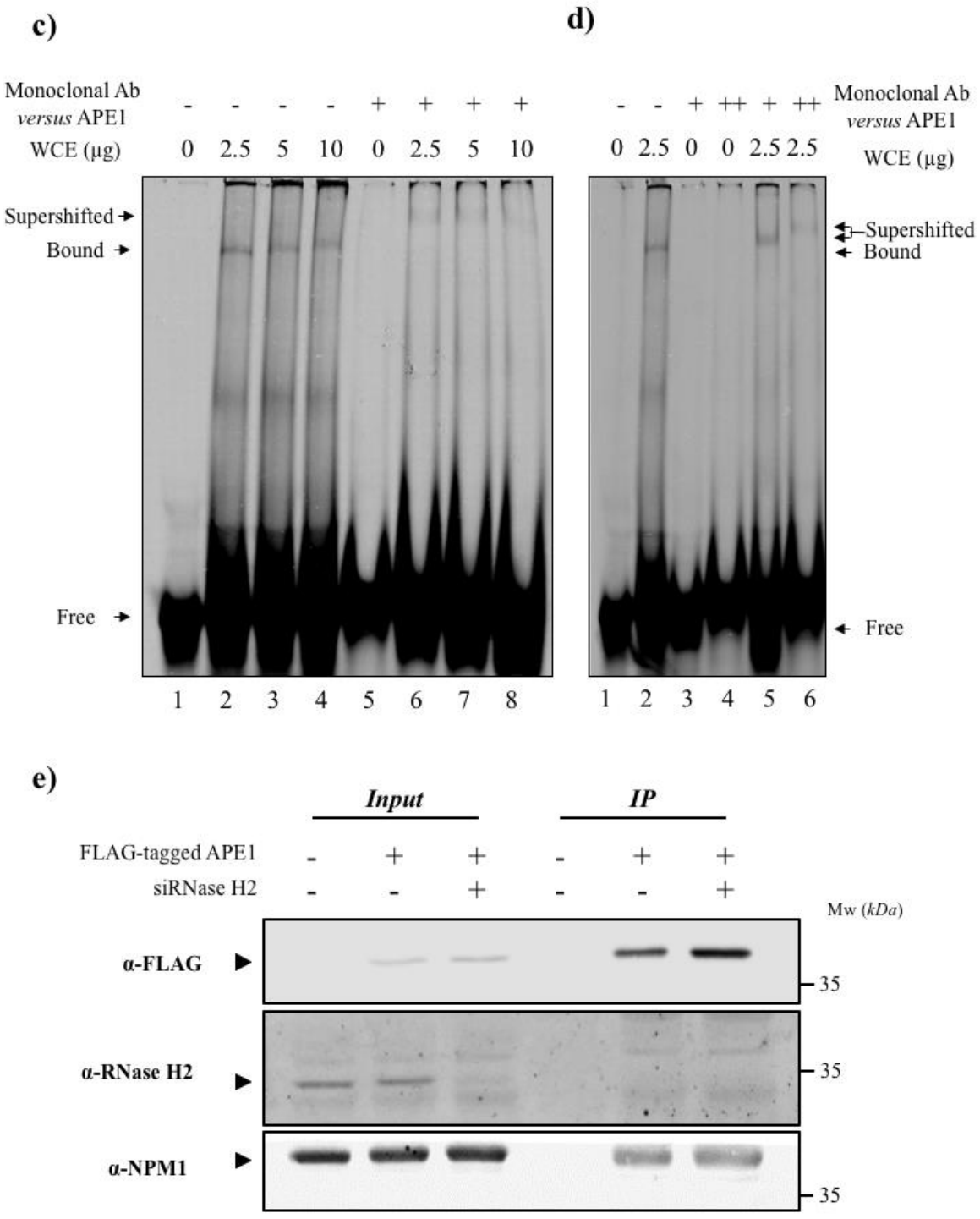


Figure B.7APE1 is part of the protein complex recognizing r8oxoG from cell extracts, as evaluated through *supershift* EMSA analysis. (**a**) HeLa cells were transfected with siRNAs directed versus RNase H2 and APE1 proteins and compared to the Scramble control. Whole cell extracts were separated onto 12% SDS–PAGE, and Western blot analysis was performed by using an anti-APE1 antibody and an anti-RNase H2 antibody.

Tubulin was used as loading control. The molecular weight (MW) expressed in kilodaltons (*kDa*) is shown on the right of each panel. (**b**) Representative native EMSA polyacrylamide gel on ds_r8oxoG:dC oligonucleotide incubated with 2.5 µg of whole cell extracts from untreated HeLa cells (*lane 2*) and HeLa cells transfected with control siRNA (*lane 3*), silenced for APE1 (*lane 4*), RNase H2 (*lane 5*) and APE1/RNase H2 (*lane 6*) for 30 minutes at 37°C. Lane 1 corresponds to r8oxoG-containing oligonucleotide alone without any cell extract. On the right side of the gel, an arrow denotes the retarded complex between cell extract and the probe (*left*). Histogram shows the difference of binding between each condition (*right*). (**c**) Representative native polyacrylamide gel of ds_r8oxoG:dC oligonucleotide incubated at 37°C for 30 minutes with the indicated amounts of whole cell extracts (WCE) without or with a pre-incubation with the specific monoclonal antibody recognizing APE1 protein, at 4°C for 1 hour, before gel separation. Bound complex to the oligonucleotide is indicated by an arrow named 'Bound', whereas the *supershifted* complex, obtained after incubation with the specific antibody, is indicated by an arrow named 'Supershifted'. (**d**) Representative native polyacrylamide gel of ds_r8oxoG:dC oligonucleotide incubated at 37°C for 30 minutes with the indicated amounts of whole cell extracts (WCE) without or with a pre-incubation with increasing amounts of the specific monoclonal antibody recognizing APE1 protein, at 4°C for 1 hour, before gel separation. Bound complex to the oligonucleotide is indicated by an arrow named 'Bound', whereas the *supershifted* complex, obtained after incubation with the specific antibody, is indicated by an arrow named 'Supershifted'. (**e**) Western blot analysis performed on total cell extracts (Input) (*left*) and on immunoprecipitated material (IP) (*right*) from HeLa cells previously transfected with FLAG-tagged APE1 expressing plasmids. Silencing of RNase H2 was performed to demonstrate the specificity of the recognized band by the anti-RNAse H2 antibody used. Whole cell extracts were immunoprecipitated with a with α-FLAG resin. Co-immunoprecipitation with NPM1 was used as positive control as already described previously [67].

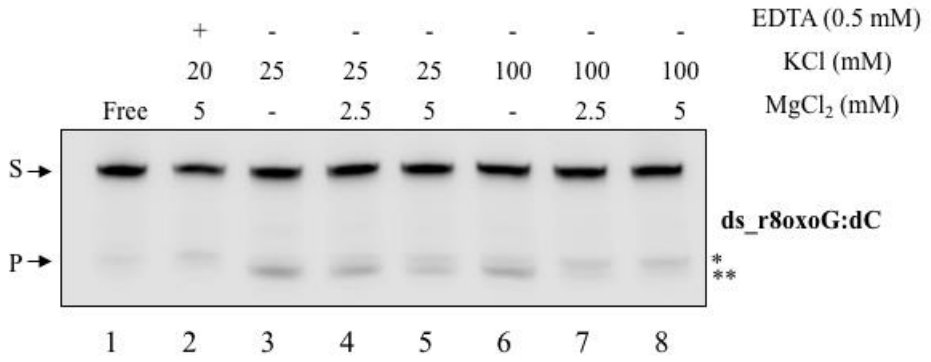


Figure B.8 APE1 possesses NIR activity on r8oxoG-containing oligonucleotide under specific experimental conditions. Comparison of the reaction conditions used in this study with that by Sassa et al. Representative denaturing polyacrylamide gel of ds_r8oxoG:dC oligonucleotide incision by APE1 under different buffer conditions. On top of the panel MgCl2 and KCl concentrations (mM) are indicated for each experimental condition. The presence of EDTA is indicated where is the case. Lane 2 reproduces buffer conditions from Sassa et al. [154]. S indicates the substrate position while P indicates the product position (left). Two asterisks indicate the position of the two products (right).

APPENDIX C. Supplementary information for Chapter 5

Table C.1 All strains used in this study

Strain	Relevant genotype	Source
E134 (KK-44)	*MATα ade5-1 lys2-14A trp1-289 his7-2 leu2-3,112 ura3-52*	Koh *et al.*, 2015 [70]
KK-100	*KK-44 rnh201Δ::hygMX4*	Koh *et al.*, 2015 [70]
KK-172	KK-44 *rnh1Δ::kanMX4*	this study
SB-286	KK-44 *rnh202-FF346,347AA*	this study
BY4742 (KK-2)	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3* Δ0	Storici *et al.*, 2001 [183]
SB-305	KK-2 *rnh201Δ::hygMX4*	this study
SB-285	KK-2 *rnh202-FF346,347AA*	this study
SB-311	KK-2 *rnh201-P45D Y219A*	this study
YFP17 (SB-288)	*Δho Δhml::ADE1 mataΔ::hisG Δhmr::ADE1 ade1 leu2-3,1122 lys5 trp1::hisG ura3-52 ade3::GAL::HO*	Keskin *et al.*, 2015 [184]
SB-293	SB-293 *rnh201Δ::hygMX4*	this study
DG2204 (HK-692)	*MATa LYS+ trp1 ADE+ LEU+ ura3 his3-Δ200 hisG Gal+ Spo+ Tyless*	Garfinkel *et* al., 2005 [185]
HK-705	HK-692 *rnh201Δ::hygMX4*	this study
JZ105 (KK-154)	*Mat1M Δmat2,3::LEU2 ade6-210 leu1-32 ura4-D18 his2*	Vengrova *et al.*, 2004 [42]
HK-983	KK-154 *rnh201Δ::kanMX4*	this study

All the yeast strains used in ribose-seq library construction and their corresponding genotype.

Table C.2 List of strains used and the corresponding restriction enzymes used for fragmentation of genomic DNA.

Strain	FS #	Enzymes used
E134 (KK-44)	FS 93	DraI, EcoRV, SspI (DES)
KK-100	FS 95	DES
	FS 115	RsaI, HaeIII (RH)
	FS117	DraI, EcoRV, SspI, AluI (DESA)
	FS 118	DESA
	FS 141	DESA
	FS 156	DESA
BY4742 (KK-2)	FS 106	DES
SB-305	FS 138	DESA
SB-285	FS 150	DES
SB-311	FS 153	DES
YFP17 (SB-288)	FS 107	DES
SB-293	FS 140	RH
DG2204 (HK-692)	FS 129	DESA
HK-705	FS 130	DESA
JZ105 (KK-154)	FS 94	DES
HK-983	FS 97	DES

Ribose-seq libraries were constructed by fragmentation of genomic DNA using restriction endonucleases. Three sets of enzymes were used: (i) RsaI, HaeIII (RH), (ii) DraI, EcoRV, SspI (DES), and (iii) DraI, EcoRV, SspI and AluI (DESA). The library names and the corresponding enzymes used for their construction are indicated. The enzymes were purchased from New England Biolabs.

Table C.3 List of oligonucleotides used in this study

Name	Size	Sequence
202PIP.F	86	5’AAAACCAAAAGTAGCCATAGGAAAAGGGGCCATTGATGGA**GCTGCT**AAACGTAAGTAGCTAGTATCATAATTAAACAGCAATTTGA
202PIP.R	86	5’TCAAATTGCTGTTTAATTATGATACTAGCTACTTACGTTT**AGCAGC**TCCATCAATGGCCCCTTTTCCTATGGCTACTTTTGGTTTT
RNH P45D.60	60	5’AATGGGTATCGATGAAGCTGGCAGAGGG**GA**CGTATTAGGGCCAATGGTCTACGCAGTAG
RNH Y219A.60	60	5’GGGATCCCGATGAGATCCTGGGTTCTGGA**GC**CCCCTCCGACCCGAAGACAGTCGCATGGC
Adaptor.L1 *	65	5’ P-NNC CGN NNN NNA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTT GAT GAT AGA TCC GTG TCG CAA CT
Adaptor.L2 *	65	5’ P-NNT GAN NNN NNA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTT GAT GAT AGA TCC GTG TCG CAA CT
Adaptor.S *	25	5’ P-GTT GCG ACA CGG ATC TAT CAA CAC T -Am
PCR.1	54	5’ GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TTG ATA GAT CCG TGT CGC AAC
PCR.2	20	5’ ACA CTC TTT CCC TAC ACG AC
PCR.701	53	5’ CAA GCA GAA GAC GGC ATA CGA GAT **CGA GTA AT**G TGA CTG GAG TTC AGA CGT GT
PCR.702	53	5’ CAA GCA GAA GAC GGC ATA CGA GAT **TCT CCG GA**G TGA CTG GAG TTC AGA CGT GT
PCR.501	57	5’ AAT GAT ACG GCG ACC ACC GAG ATC TAC AC**T ATA GCC T**AC ACT CTT TCC CTA CAC GAC
PCR.502	57	5’ AAT GAT ACG GCG ACC ACC GAG ATC TAC AC**A TAG AGG C**AC ACT CTT TCC CTA CAC GAC
PCR.503	57	5’ AAT GAT ACG GCG ACC ACC GAG ATC TAC AC**C CTA TCC T**AC ACT CTT TCC CTA CAC GAC

PCR.504	57	5’ AAT GAT ACG GCG ACC ACC GAG ATC TAC AC**G GCT CTG A**AC ACT CTT TCC CTA CAC GAC
PCR.505	57	5’ AAT GAT ACG GCG ACC ACC GAG ATC TAC AC**A GGC GAA G**AC ACT CTT TCC CTA CAC GAC
PCR.506	57	5’ AAT GAT ACG GCG ACC ACC GAG ATC TAC AC**T AAT CTT A**AC ACT CTT TCC CTA CAC GAC
PCR.507	57	5’ AAT GAT ACG GCG ACC ACC GAG ATC TAC AC**C AGG ACG T**AC ACT CTT TCC CTA CAC GAC
PCR.508	57	5’ AAT GAT ACG GCG ACC ACC GAG ATC TAC AC**G TAC TGA C**AC ACT CTT TCC CTA CAC GAC

Name, length, and sequence of oligonucleotides used in this study are described. The bold and underlined letters in oligonucleotides (202.PIP.F, 202.PIP.R, RNH P45D.60, and RNH Y219A.60) indicate the specific mutations that were introduced in RNH202 or RNH201 region. All bold letters in the PCR primers indicate the specific sequence of index used in sequencing. ‘P and Am’ indicate end modifications of phosphate and amino groups respectively. All oligonucleotides were desalted, except for those marked in *, which were HPLC purified. 202.PIP.F, 202.PIP.R, RNH P45D.60, and RNH Y219A.60 were synthesized by Eurofins Technologies, and all other oligonucleotides were synthesized by IDT.

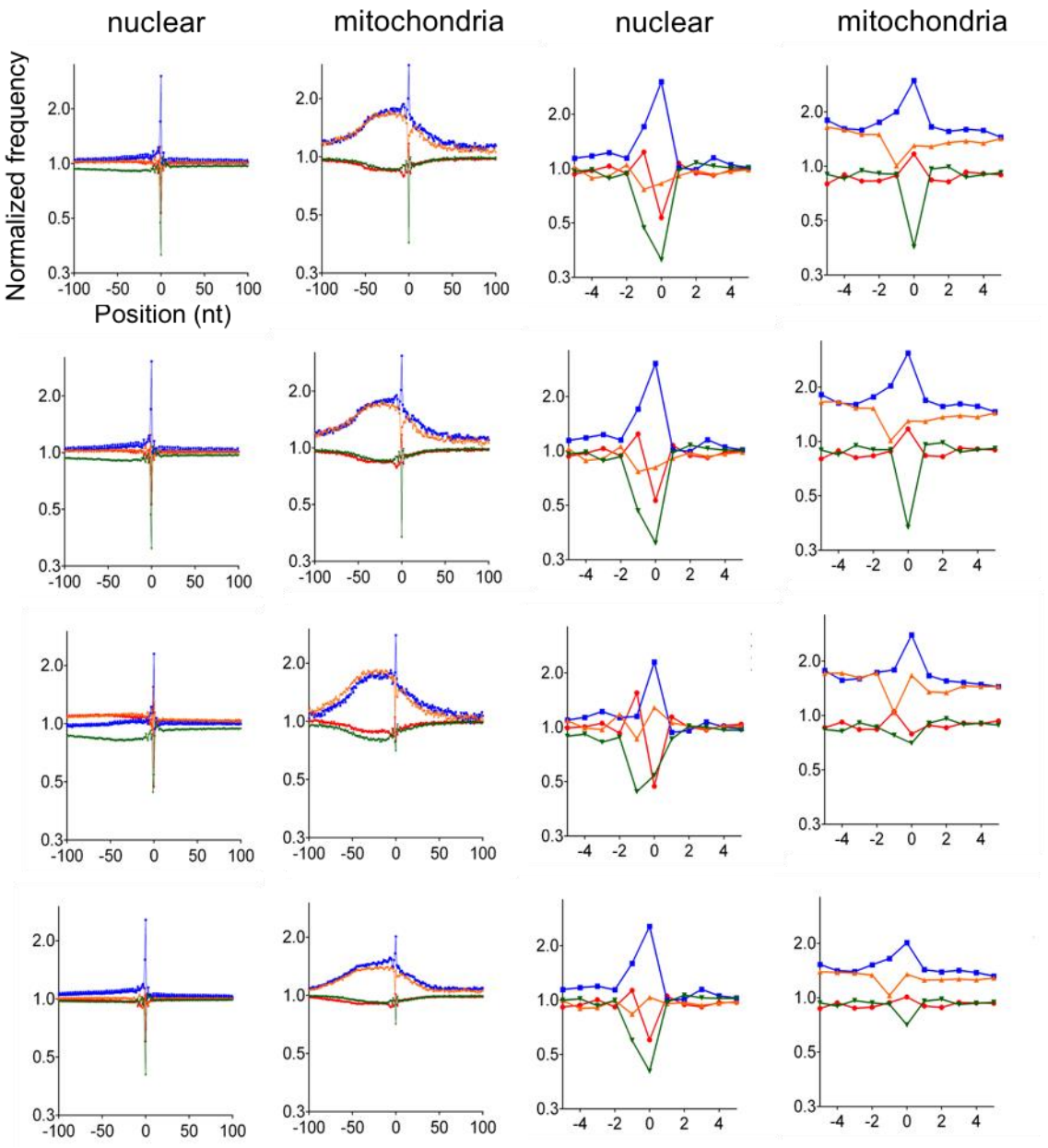


Figure C.1 Identity and sequence context of rNMP incorporation in *S. cerevisiae rnh201*Δ genome. Shown here are normalized nucleotide plots relative to mapped positions of sequences from ribose-seq library of *rnh201Δ* cells (KK-100). Four independent libraries were prepared from *rnh201Δ* cells (KK-100), FS117 (row 1), FS118 (row 2), FS141 (row 3), and FS156 (row 4). All these libraries were prepared by using DraI, EcoRV, SspI, and AluI restriction enzymes for fragmenting their genomic DNA.

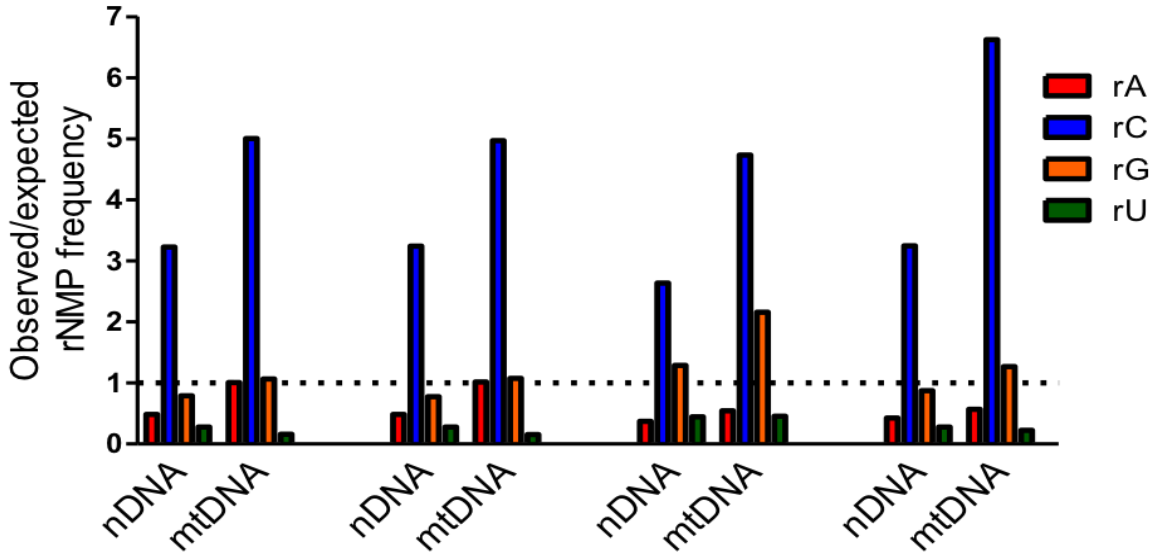


Figure C.2 rNMP composition in *S. cerevisiae rnh201*Δ cells. The observed rNMP frequencies from **Figure C.1** were divided by the expected proportion of bases shown in **Table C.4** for four independent libraries of *S. cerevisiae rnh201*Δ (KK-100) genome. Shown here are bar graphs from one library.

REFERENCES

1. Sparks, J.L., et al., *RNase H2-initiated ribonucleotide excision repair.* Mol Cell, 2012. **47**(6): p. 980-6.

2. Caldecott, K.W., *Molecular biology. Ribose--an internal threat to DNA.* Science, 2014. **343**(6168): p. 260-1.

3. Peterson, C.L. and J. Cote, *Cellular machineries for chromosomal DNA repair.* Genes Dev, 2004. **18**(6): p. 602-16.

4. Bhattacharjee, S. and S. Nandi, *Choices have consequences: the nexus between DNA repair pathways and genomic instability in cancer.* Clin Transl Med, 2016. **5**(1): p. 45.

5. Lindahl, T., *Instability and decay of the primary structure of DNA.* Nature, 1993. **362**(6422): p. 709-15.

6. Williams, J.S. and T.A. Kunkel, *Ribonucleotides in DNA: origins, repair and consequences.* DNA Repair (Amst), 2014. **19**: p. 27-37.

7. Cadet, J., E. Sage, and T. Douki, *Ultraviolet radiation-mediated damage to cellular DNA.* Mutat Res, 2005. **571**(1-2): p. 3-17.

8. Bauer, N.C., A.H. Corbett, and P.W. Doetsch, *The current state of eukaryotic DNA base damage and repair.* Nucleic Acids Res, 2015. **43**(21): p. 10083-101.

9. Iyer, R.R., et al., *DNA mismatch repair: functions and mechanisms.* Chem Rev, 2006. **106**(2): p. 302-23.

10. Hoeijmakers, J.H., *Genome maintenance mechanisms for preventing cancer.* Nature, 2001. **411**(6835): p. 366-74.

11. Madabhushi, R., L. Pan, and L.H. Tsai, *DNA damage and its links to neurodegeneration.* Neuron, 2014. **83**(2): p. 266-282.

12. Nick McElhinny, S.A., et al., *Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases.* Proc Natl Acad Sci U S A, 2010. **107**(11): p. 4949-54.

13. Traut, T.W., *Physiological concentrations of purines and pyrimidines.* Mol Cell Biochem, 1994. **140**(1): p. 1-22.

14. Joyce, C.M., *Choosing the right sugar: how polymerases select a nucleotide substrate.* Proc Natl Acad Sci U S A, 1997. **94**(5): p. 1619-22.

15. Ide, H., et al., *Misincorporation of ribonucleotides by DNA polymerase during in vitro DNA replication.* Nucleic Acids Symp Ser, 1993(29): p. 133-4.

16. McDonald, J.P., et al., *Mechanisms employed by Escherichia coli to prevent ribonucleotide incorporation into genomic DNA by Pol V.* PLoS Genet, 2012. **8**(11): p. e1003030.

17. Zhu, H. and S. Shuman, *Bacterial nonhomologous end joining ligases preferentially seal breaks with a 3'-OH monoribonucleotide.* J Biol Chem, 2008. **283**(13): p. 8331-9.

18. Clausen, A.R., et al., *Ribonucleotide incorporation, proofreading and bypass by human DNA polymerase delta.* DNA Repair (Amst), 2013. **12**(2): p. 121-7.

19. Su, Y., M. Egli, and F.P. Guengerich, *Mechanism of Ribonucleotide Incorporation by Human DNA Polymerase eta.* J Biol Chem, 2016. **291**(8): p. 3747-56.

20. Goksenin, A.Y., et al., *Human DNA polymerase epsilon is able to efficiently extend from multiple consecutive ribonucleotides.* J Biol Chem, 2012. **287**(51): p. 42675-84.

21. Burgers, P.M., *Polymerase dynamics at the eukaryotic DNA replication fork.* J Biol Chem, 2009. **284**(7): p. 4041-5.

22. Rumbaugh, J.A., et al., *Creation and removal of embedded ribonucleotides in chromosomal DNA during mammalian Okazaki fragment processing.* J Biol Chem, 1997. **272**(36): p. 22591-9.

23. Wiseman, H. and B. Halliwell, *Damage to DNA by reactive oxygen and nitrogen species: role in inflammatory disease and progression to cancer.* Biochem J, 1996. **313 (Pt 1)**: p. 17-29.

24. Randerath, K., et al., *Formation of ribonucleotides in DNA modified by oxidative damage in vitro and in vivo. Characterization by 32P-postlabeling.* Mutat Res, 1992. **275**(3-6): p. 355-66.

25. Kochetkov, N.K., Budovskii, E.I., *Reactions involving the cleavage or rearrangement of heterocyclic rings of nucleic acid bases and their derivatives* Organic Chemistry of Nucleic Acids, 1972: p. 381-423.

26. Lindahl, T. and B. Nyberg, *Rate of depurination of native deoxyribonucleic acid.* Biochemistry, 1972. **11**(19): p. 3610-8.

27. Williams, J.S., S.A. Lujan, and T.A. Kunkel, *Processing ribonucleotides incorporated during eukaryotic DNA replication.* Nat Rev Mol Cell Biol, 2016. **17**(6): p. 350-63.

28. Watt, D.L., et al., *Replication of ribonucleotide-containing DNA templates by yeast replicative polymerases.* DNA Repair (Amst), 2011. **10**(8): p. 897-902.

29. Shen, Y., et al., *Mispaired rNMPs in DNA are mutagenic and are targets of mismatch repair and RNases H.* Nat Struct Mol Biol, 2011. **19**(1): p. 98-104.

30. Shen, Y., et al., *RNA-driven genetic changes in bacteria and in human cells.* Mutat Res, 2011. **717**(1-2): p. 91-8.

31. Storici, F., et al., *RNA-templated DNA repair.* Nature, 2007. **447**(7142): p. 338-41.

32. Clausen, A.R., et al., *Structure-function analysis of ribonucleotide bypass by B family DNA replicases.* Proc Natl Acad Sci U S A, 2013. **110**(42): p. 16802-7.

33. Lazzaro, F., et al., *RNase H and postreplication repair protect cells from ribonucleotides incorporated in DNA.* Mol Cell, 2012. **45**(1): p. 99-110.

34. Pizzi, S., et al., *Reduction of hRNase H2 activity in Aicardi-Goutieres syndrome cells leads to replication stress and genome instability.* Hum Mol Genet, 2015. **24**(3): p. 649-58.

35. Williams, J.S., et al., *Topoisomerase 1-mediated removal of ribonucleotides from nascent leading-strand DNA.* Mol Cell, 2013. **49**(5): p. 1010-5.

36. DeRose, E.F., et al., *Solution structure of the Dickerson DNA dodecamer containing a single ribonucleotide.* Biochemistry, 2012. **51**(12): p. 2407-16.

37. Chiu, H.C., et al., *RNA intrusions change DNA elastic properties and structure.* Nanoscale, 2014. **6**(17): p. 10009-17.

38. Hovatter, K.R. and H.G. Martinson, *Ribonucleotide-induced helical alteration in DNA prevents nucleosome formation.* Proc Natl Acad Sci U S A, 1987. **84**(5): p. 1162-6.

39. Dunn, K. and J.D. Griffith, *The presence of RNA in a double helix inhibits its interaction with histone protein.* Nucleic Acids Res, 1980. **8**(3): p. 555-66.

40. Ghodgaonkar, M.M., et al., *Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair.* Mol Cell, 2013. **50**(3): p. 323-32.

41. Lujan, S.A., et al., *Ribonucleotides are signals for mismatch repair of leading-strand replication errors.* Mol Cell, 2013. **50**(3): p. 437-43.

42. Vengrova, S. and J.Z. Dalgaard, *The wild-type Schizosaccharomyces pombe mat1 imprint consists of two ribonucleotides.* EMBO Rep, 2006. **7**(1): p. 59-65.

43. Ohtani, N., et al., *Molecular diversities of RNases H.* J Biosci Bioeng, 1999. **88**(1): p. 12-9.

44. Cerritelli, S.M. and R.J. Crouch, *Ribonuclease H: the enzymes in eukaryotes.* Febs j, 2009. **276**(6): p. 1494-505.

45. Rychlik, M.P., et al., *Crystal structures of RNase H2 in complex with nucleic acid reveal the mechanism of RNA-DNA junction recognition and cleavage.* Mol Cell, 2010. **40**(4): p. 658-70.

46. Bubeck, D., et al., *PCNA directs type 2 RNase H activity on DNA replication and repair substrates.* Nucleic Acids Res, 2011. **39**(9): p. 3652-66.

47. Chon, H., et al., *Contributions of the two accessory subunits, RNASEH2B and RNASEH2C, to the activity and properties of the human RNase H2 complex.* Nucleic Acids Res, 2009. **37**(1): p. 96-110.

48. Kind, B., et al., *Altered spatio-temporal dynamics of RNase H2 complex assembly at replication and repair sites in Aicardi-Goutieres syndrome.* Hum Mol Genet, 2014. **23**(22): p. 5950-60.

49. Aguilera, A. and H.L. Klein, *Genetic control of intrachromosomal recombination in Saccharomyces cerevisiae. I. Isolation and genetic characterization of hyper-recombination mutations.* Genetics, 1988. **119**(4): p. 779-90.

50. Potenski, C.J., et al., *Avoidance of ribonucleotide-induced mutations by RNase H2 and Srs2-Exo1 mechanisms.* Nature, 2014. **511**(7508): p. 251-4.

51. Allen-Soltero, S., et al., *A saccharomyces cerevisiae RNase H2 interaction network functions to suppress genome instability.* Mol Cell Biol, 2014. **34**(8): p. 1521-34.

52. Wahba, L., et al., *RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability.* Mol Cell, 2011. **44**(6): p. 978-88.

53. Reijns, M.A., et al., *Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development.* Cell, 2012. **149**(5): p. 1008-22.

54. Crow, Y.J., et al., *Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutieres syndrome and mimic congenital viral brain infection.* Nat Genet, 2006. **38**(8): p. 910-6.

55. Gunther, C., et al., *Defective removal of ribonucleotides from DNA promotes systemic autoimmunity.* J Clin Invest, 2015. **125**(1): p. 413-24.

56. Kim, N., et al., *Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I.* Science, 2011. **332**(6037): p. 1561-4.

57. Sekiguchi, J. and S. Shuman, *Site-specific ribonuclease activity of eukaryotic DNA topoisomerase I.* Mol Cell, 1997. **1**(1): p. 89-97.

58. Pommier, Y., et al., *Roles of eukaryotic topoisomerases in transcription, replication and genomic stability.* Nat Rev Mol Cell Biol, 2016. **17**(11): p. 703-721.

59. Huang, S.Y., S. Ghosh, and Y. Pommier, *Topoisomerase I alone is sufficient to produce short DNA deletions and can also reverse nicks at ribonucleotide sites.* J Biol Chem, 2015. **290**(22): p. 14068-76.

60. Prakash, S. and L. Prakash, *Nucleotide excision repair in yeast.* Mutat Res, 2000. **451**(1-2): p. 13-24.

61. Lindsey-Boltz, L.A., et al., *Analysis of Ribonucleotide Removal from DNA by Human Nucleotide Excision Repair.* J Biol Chem, 2015. **290**(50): p. 29801-7.

62. Shcherbakova, P.V., et al., *Unique error signature of the four-subunit yeast DNA polymerase epsilon.* J Biol Chem, 2003. **278**(44): p. 43770-80.

63. Williams, J.S., et al., *Proofreading of ribonucleotides inserted into DNA by yeast DNA polymerase varepsilon.* DNA Repair (Amst), 2012. **11**(8): p. 649-56.

64. Dianov, G.L., et al., *Base excision repair in nuclear and mitochondrial DNA.* Prog Nucleic Acid Res Mol Biol, 2001. **68**: p. 285-97.

65. Krokan, H.E. and M. Bjoras, *Base excision repair.* Cold Spring Harb Perspect Biol, 2013. **5**(4): p. a012583.

66. Tell, G., D.M. Wilson, 3rd, and C.H. Lee, *Intrusion of a DNA repair protein in the RNome world: is this the beginning of a new era?* Mol Cell Biol, 2010. **30**(2): p. 366-71.

67. Vascotto, C., et al., *APE1/Ref-1 interacts with NPM1 within nucleoli and plays a role in the rRNA quality control process.* Mol Cell Biol, 2009. **29**(7): p. 1834-54.

68. Clausen, A.R., et al., *Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation.* Nat Struct Mol Biol, 2015. **22**(3): p. 185-91.

69. Daigaku, Y., et al., *A global profile of replicative polymerase usage.* Nat Struct Mol Biol, 2015. **22**(3): p. 192-198.

70. Koh, K.D., et al., *Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA.* Nat Methods, 2015. **12**(3): p. 251-7, 3 p following 257.

71. Reijns, M.A.M., et al., *Lagging-strand replication shapes the mutational landscape of the genome.* Nature, 2015. **518**(7540): p. 502-506.

72. Grossman, L.I., R. Watson, and J. Vinograd, *The presence of ribonucleotides in mature closed-circular mitochondrial DNA.* Proc Natl Acad Sci U S A, 1973. **70**(12): p. 3339-43.

73. Vengrova, S. and J.Z. Dalgaard, *The wild-type Schizosaccharomyces pombe mat1 imprint consists of two ribonucleotides.* EMBO Rep, 2006. **7**: p. 59-65.

74. Potenski, C.J. and H.L. Klein, *How the misincorporation of ribonucleotides into genomic DNA can be both harmful and helpful to cells.* Nucleic Acids Res, 2014.

75. Kasiviswanathan, R. and W.C. Copeland, *Ribonucleotide discrimination and reverse transcription by the human mitochondrial DNA polymerase.* J Biol Chem, 2011. **286**(36): p. 31490-500.

76. Yao, N.Y., et al., *Cost of rNTP/dNTP pool imbalance at the replication fork.* Proc Natl Acad Sci U S A, 2013. **110**(32): p. 12942-7.

77. Kim, N., et al., *Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I.* Science, 2011. **332**(6037): p. 1561-4.

78. Potenski, C.J., et al., *Avoidance of ribonucleotide-induced mutations by RNase H2 and Srs2-Exo1 mechanisms.* Nature, 2014.

79. Cho, J.E., et al., *Two distinct mechanisms of Topoisomerase 1-dependent mutagenesis in yeast.* DNA Repair (Amst), 2013. **12**(3): p. 205-11.

80. Clark, A.B., et al., *Functional analysis of human MutSalpha and MutSbeta complexes in yeast.* Nucleic Acids Res, 1999. **27**(3): p. 736-42.

81. Morrison, A., et al., *Eukaryotic DNA polymerase amino acid sequence required for 3'----5' exonuclease activity.* Proc Natl Acad Sci U S A, 1991. **88**(21): p. 9473-7.

82. Jin, Y.H., et al., *The 3'-->5' exonuclease of DNA polymerase delta can substitute for the 5' flap endonuclease Rad27/Fen1 in processing Okazaki fragments and preventing genome instability.* Proc Natl Acad Sci U S A, 2001. **98**(9): p. 5122-7.

83. Pursell, Z.F., et al., *Yeast DNA polymerase epsilon participates in leading-strand DNA replication.* Science, 2007. **317**(5834): p. 127-30.

84. Schutz, K., J.R. Hesselberth, and S. Fields, *Capture and sequence analysis of RNAs with terminal 2',3'-cyclic phosphates.* RNA, 2010. **16**(3): p. 621-31.

85. Kokoska, R.J., S.D. McCulloch, and T.A. Kunkel, *The efficiency and specificity of apurinic/apyrimidinic site bypass by human DNA polymerase eta and Sulfolobus solfataricus Dpo4.* J Biol Chem, 2003. **278**(50): p. 50537-45.

86. Kuchta, R.D. and G. Stengel, *Mechanism and evolution of DNA primases.* Biochim Biophys Acta, 2010. **1804**(5): p. 1180-9.

87. Smith, C.W.J., *RNA-protein interactions : a practical approach*. Practical approach series. 1998, Oxford ; New York: Oxford University Press. xxv, 341 p.

88. Lhomme, J., J.F. Constant, and M. Demeunynck, *Abasic DNA structure, reactivity, and recognition.* Biopolymers, 1999. **52**(2): p. 65-83.

89. Bailly, V. and W.G. Verly, *Possible roles of beta-elimination and delta-elimination reactions in the repair of DNA containing AP (apurinic/apyrimidinic) sites in mammalian cells.* Biochem J, 1988. **253**(2): p. 553-9.

90. Barrett, T., et al., *NCBI GEO: archive for functional genomics data sets--update.* Nucleic Acids Res, 2013. **41**(Database issue): p. D991-5.

91. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-2.

92. Dale, R.K., B.S. Pedersen, and A.R. Quinlan, *Pybedtools: a flexible Python library for manipulating genomic datasets and annotations.* Bioinformatics, 2011. **27**(24): p. 3423-4.

93. Siow, C.C., et al., *OriDB, the DNA replication origin database updated and extended.* Nucleic Acids Res, 2012. **40**(Database issue): p. D682-6.

94. Yabuki, N., H. Terashima, and K. Kitada, *Mapping of early firing origins on a replication profile of budding yeast.* Genes to cells : devoted to molecular & cellular mechanisms, 2002. **7**(8): p. 781-9.

95. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.

96. Sokal, R.R. and F.J. Rohlf, *Biometry : the principles and practice of statistics in biological research*. 3rd ed. 1995, New York: W.H. Freeman. xix, 887 p.

97. Remus, B.S. and S. Shuman, *Distinctive kinetics and substrate specificities of plant and fungal tRNA ligases.* RNA, 2014. **20**(4): p. 462-73.

98. Cooper, D.A., et al., *Ribonuclease L and metal-ion-independent endoribonuclease cleavage sites in host and viral RNAs.* Nucleic Acids Res, 2014. **42**(8): p. 5202-16.

99. Krokan, H.E., F. Drablos, and G. Slupphaug, *Uracil in DNA--occurrence, consequences and repair.* Oncogene, 2002. **21**(58): p. 8935-48.

100. Lindahl, T., et al., *DNA N-glycosidases: properties of uracil-DNA glycosidase from Escherichia coli.* J Biol Chem, 1977. **252**(10): p. 3286-94.

101. Fasullo, M., et al., *Elevated dNTP levels suppress hyper-recombination in Saccharomyces cerevisiae S-phase checkpoint mutants.* Nucleic Acids Res, 2010. **38**(4): p. 1195-203.

102. Foury, F., et al., *The complete sequence of the mitochondrial genome of Saccharomyces cerevisiae.* FEBS Lett, 1998. **440**(3): p. 325-31.

103. Gerhold, J.M., et al., *Strand invasion structures in the inverted repeat of Candida albicans mitochondrial DNA reveal a role for homologous recombination in replication.* Mol Cell, 2010. **39**(6): p. 851-61.

104. Moraes, C.T., *What regulates mitochondrial DNA copy number in animal cells?* Trends Genet, 2001. **17**(4): p. 199-205.

105. Szostak, J.W. and R. Wu, *Unequal crossing over in the ribosomal DNA of Saccharomyces cerevisiae.* Nature, 1980. **284**(5755): p. 426-30.

106. Hani, J. and H. Feldmann, *tRNA genes and retroelements in the yeast genome.* Nucleic Acids Res, 1998. **26**(3): p. 689-96.

107. Mieczkowski, P.A., F.J. Lemoine, and T.D. Petes, *Recombination between retrotransposons as a source of chromosome rearrangements in the yeast Saccharomyces cerevisiae.* DNA Repair (Amst), 2006. **5**(9-10): p. 1010-20.

108. El Hage, A., et al., *Genome-Wide Distribution of RNA-DNA Hybrids Identifies RNase H Targets in tRNA Genes, Retrotransposons and Mitochondria.* PLoS Genet, 2014. **10**(10): p. e1004716.

109. Crespan, E., et al., *Impact of ribonucleotide incorporation by DNA polymerases beta and lambda on oxidative base excision repair.* Nat Commun, 2016. **7**: p. 10805.

110. Koh, K.D., et al., *Measuring the elasticity of ribonucleotide(s)-containing DNA molecules using AFM.* Methods Mol Biol, 2015. **1297**: p. 43-57.

111. Evich, M., et al., *Structural Impact of Single Ribonucleotide Residues in DNA.* Chembiochem, 2016. **17**(20): p. 1968-1977.

112. Crow, Y.J. and N. Manel, *Aicardi-Goutieres syndrome and the type I interferonopathies.* Nat Rev Immunol, 2015. **15**(7): p. 429-40.

113. Rice, G., et al., *Clinical and molecular phenotype of Aicardi-Goutieres syndrome.* Am J Hum Genet, 2007. **81**(4): p. 713-25.

114. Brzostek-Racine, S., et al., *The DNA damage response induces IFN.* J Immunol, 2011. **187**(10): p. 5336-45.

115. Vaisman, A., et al., *Removal of misincorporated ribonucleotides from prokaryotic genomes: an unexpected role for nucleotide excision repair.* PLoS Genet, 2013. **9**(11): p. e1003878.

116. Sykora, P., D.M. Wilson, 3rd, and V.A. Bohr, *Repair of persistent strand breaks in the mitochondrial genome.* Mech Ageing Dev, 2012. **133**(4): p. 169-75.

117. Farrington, S.M., et al., *Germline susceptibility to colorectal cancer due to base-excision repair gene defects.* Am J Hum Genet, 2005. **77**(1): p. 112-9.

118. Sastre-Moreno, G., et al., *ATP insertion opposite 8-oxo-deoxyguanosine by Pol4 mediates error-free tolerance in Schizosaccharomyces pombe.* Nucleic Acids Res, 2014. **42**(15): p. 9821-37.

119. Ordonez, H. and S. Shuman, *Mycobacterium smegmatis DinB2 misincorporates deoxyribonucleotides and ribonucleotides during templated synthesis and lesion bypass.* Nucleic Acids Res, 2014. **42**(20): p. 12722-34.

120. Cilli, P., et al., *Formation and Repair of Mismatches Containing Ribonucleotides and Oxidized Bases at Repeated DNA Sequences.* J Biol Chem, 2015. **290**(43): p. 26259-69.

121. Barnes, T., et al., *Identification of Apurinic/apyrimidinic endonuclease 1 (APE1) as the endoribonuclease that cleaves c-myc mRNA.* Nucleic Acids Res, 2009. **37**(12): p. 3946-58.

122. Berquist, B.R., D.R. McNeill, and D.M. Wilson, 3rd, *Characterization of abasic endonuclease activity of human Ape1 on alternative substrates, as well as effects of ATP and sequence context on AP site incision.* J Mol Biol, 2008. **379**(1): p. 17-27.

123. Li, M. and D.M. Wilson, 3rd, *Human apurinic/apyrimidinic endonuclease 1.* Antioxid Redox Signal, 2014. **20**(4): p. 678-707.

124. Chohan, M., et al., *Human apurinic/apyrimidinic endonuclease 1 (APE1) has 3' RNA phosphatase and 3' exoribonuclease activities.* J Mol Biol, 2015. **427**(2): p. 298-311.

125. Gros, L., et al., *The major human AP endonuclease (Ape1) is involved in the nucleotide incision repair pathway.* Nucleic Acids Res, 2004. **32**(1): p. 73-81.

126. Daviet, S., et al., *Major oxidative products of cytosine are substrates for the nucleotide incision repair pathway.* DNA Repair (Amst), 2007. **6**(1): p. 8-18.

127. Mazouzi, A., et al., *Insight into mechanisms of 3'-5' exonuclease activity and removal of bulky 8,5'-cyclopurine adducts by apurinic/apyrimidinic endonucleases.* Proc Natl Acad Sci U S A, 2013. **110**(33): p. E3071-80.

128. Gasparutto, D., et al., *Chemical synthesis of a biologically active natural tRNA with its minor bases.* Nucleic Acids Res, 1992. **20**(19): p. 5159-66.

129. Audebert, M., J.P. Radicella, and M. Dizdaroglu, *Effect of single mutations in the OGG1 gene found in human tumors on the substrate specificity of the Ogg1 protein.* Nucleic Acids Res, 2000. **28**(14): p. 2672-8.

130. Fantini, D., et al., *Critical lysine residues within the overlooked N-terminal domain of human APE1 regulate its biological functions.* Nucleic Acids Res, 2010. **38**(22): p. 8239-56.

131. Erzberger, J.P. and D.M. Wilson, 3rd, *The role of Mg2+ and specific amino acid residues in the catalytic reaction of the major human abasic endonuclease: new insights from EDTA-resistant incision of acyclic abasic site analogs and site-directed mutagenesis.* J Mol Biol, 1999. **290**(2): p. 447-57.

132. Chon, H., et al., *RNase H2 roles in genome integrity revealed by unlinking its activities.* Nucleic Acids Res, 2013. **41**(5): p. 3130-43.

133. Wilson, D.M., 3rd, *Ape1 abasic endonuclease activity is regulated by magnesium and potassium concentrations and is robust on alternative DNA structures.* J Mol Biol, 2005. **345**(5): p. 1003-14.

134. Poletto, M., et al., *Inhibitors of the apurinic/apyrimidinic endonuclease 1 (APE1)/nucleophosmin (NPM1) interaction that display anti-tumor properties.* Mol Carcinog, 2016. **55**(5): p. 688-704.

135. Beernink, P.T., et al., *Two divalent metal ions in the active site of a new crystal form of human apurinic/apyrimidinic endonuclease, Ape1: implications for the catalytic mechanism.* J Mol Biol, 2001. **307**(4): p. 1023-34.

136. Izumi, T., et al., *Two essential but distinct functions of the mammalian abasic endonuclease.* Proc Natl Acad Sci U S A, 2005. **102**(16): p. 5739-43.

137. Rai, G., et al., *Small Molecule Inhibitors of the Human Apurinic/apyrimidinic Endonuclease 1 (APE1)*, in *Probe Reports from the NIH Molecular Libraries Program*. 2010, National Center for Biotechnology Information (US): Bethesda (MD).

138. Boiteux, S. and J.P. Radicella, *The human OGG1 gene: structure, functions, and its implication in the process of carcinogenesis.* Arch Biochem Biophys, 2000. **377**(1): p. 1-8.

139. Boiteux, S. and J.P. Radicella, *Base excision repair of 8-hydroxyguanine protects DNA from endogenous oxidative stress.* Biochimie, 1999. **81**(1-2): p. 59-67.

140. David, S.S., V.L. O'Shea, and S. Kundu, *Base-excision repair of oxidative DNA damage.* Nature, 2007. **447**(7147): p. 941-50.

141. Chou, K.M. and Y.C. Cheng, *The exonuclease activity of human apurinic/apyrimidinic endonuclease (APE1). Biochemical properties and inhibition by the natural dinucleotide Gp4G.* J Biol Chem, 2003. **278**(20): p. 18289-96.

142. Yasui, A., *Alternative excision repair pathways.* Cold Spring Harb Perspect Biol, 2013. **5**(6).

143. Prorok, P., et al., *Uracil in duplex DNA is a substrate for the nucleotide incision repair pathway in human cells.* Proc Natl Acad Sci U S A, 2013. **110**(39): p. E3695-703.

144. Potenski, C.J. and H.L. Klein, *How the misincorporation of ribonucleotides into genomic DNA can be both harmful and helpful to cells.* Nucleic Acids Res, 2014. **42**(16): p. 10226-34.

145. Heider, M.R., et al., *Defining the RNaseH2 enzyme-initiated ribonucleotide excision repair pathway in Archaea.* J Biol Chem, 2017. **292**(21): p. 8835-8845.

146. Hiller, B., et al., *Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity.* J Exp Med, 2012. **209**(8): p. 1419-26.

147. Rice, G.I., et al., *Synonymous mutations in RNASEH2A create cryptic splice sites impairing RNase H2 enzyme function in Aicardi-Goutieres syndrome.* Hum Mutat, 2013. **34**(8): p. 1066-70.

148. Moreira, P.I., et al., *Nucleic acid oxidation in Alzheimer disease.* Free Radic Biol Med, 2008. **44**(8): p. 1493-505.

149. Loeb, L.A. and B.D. Preston, *Mutagenesis by apurinic/apyrimidinic sites.* Annu Rev Genet, 1986. **20**: p. 201-30.

150. Jobert, L. and H. Nilsen, *Regulatory mechanisms of RNA function: emerging roles of DNA repair enzymes.* Cell Mol Life Sci, 2014. **71**(13): p. 2451-65.

151. Kochetkov, N.K., Budovskii, E.I., *Hydrolysis of N-glycosidic bonds in nucleosides, nucleotides, and their derivatives.* . Organic Chemistry of Nucleic Acids, 1972: p. 425-448.

152. Kupfer, P.A. and C.J. Leumann, *The chemical stability of abasic RNA compared to abasic DNA.* Nucleic Acids Res, 2007. **35**(1): p. 58-68.

153. Parsons, J.L., et al., *USP47 is a deubiquitylating enzyme that regulates base excision repair by controlling steady-state levels of DNA polymerase beta.* Mol Cell, 2011. **41**(5): p. 609-15.

154. Sassa, A., et al., *Impact of Ribonucleotide Backbone on Translesion Synthesis and Repair of 7,8-Dihydro-8-oxoguanine.* J Biol Chem, 2016. **291**(46): p. 24314-24323.

155. Vascotto, C., et al., *Genome-wide analysis and proteomic studies reveal APE1/Ref-1 multifunctional role in mammalian cells.* Proteomics, 2009. **9**(4): p. 1058-74.

156. Simms, C.L. and H.S. Zaher, *Quality control of chemically damaged RNA.* Cell Mol Life Sci, 2016. **73**(19): p. 3639-53.

157. Rhee, Y., M.R. Valentine, and J. Termini, *Oxidative base damage in RNA detected by reverse transcriptase.* Nucleic Acids Res, 1995. **23**(16): p. 3275-82.

158. Shan, X., Y. Chang, and C.L. Lin, *Messenger RNA oxidation is an early event preceding cell death and causes reduced protein expression.* Faseb j, 2007. **21**(11): p. 2753-64.

159. Tanaka, M., P.B. Chock, and E.R. Stadtman, *Oxidized messenger RNA induces translation errors.* Proc Natl Acad Sci U S A, 2007. **104**(1): p. 66-71.

160. Hudak, K.A., J.D. Bauman, and N.E. Tumer, *Pokeweed antiviral protein binds to the cap structure of eukaryotic mRNA and depurinates the mRNA downstream of the cap.* Rna, 2002. **8**(9): p. 1148-59.

161. Calabretta, A., P.A. Kupfer, and C.J. Leumann, *The effect of RNA base lesions on mRNA translation.* Nucleic Acids Res, 2015. **43**(9): p. 4713-20.

162. Freudenthal, B.D., et al., *Capturing snapshots of APE1 processing DNA damage.* Nat Struct Mol Biol, 2015. **22**(11): p. 924-31.

163. Schramm, V.L., *Enzymatic N-riboside scission in RNA and RNA precursors.* Curr Opin Chem Biol, 1997. **1**(3): p. 323-31.

164. Hayakawa, H., et al., *Binding capacity of human YB-1 protein for RNA containing 8-oxoguanine.* Biochemistry, 2002. **41**(42): p. 12739-44.

165. Vidal, A.E., et al., *Mechanism of stimulation of the DNA glycosylase activity of hOGG1 by the major human AP endonuclease: bypass of the AP lyase activity step.* Nucleic Acids Res, 2001. **29**(6): p. 1285-92.

166. Woo, J., et al., *Prognostic value of human apurinic/apyrimidinic endonuclease 1 (APE1) expression in breast cancer.* PLoS One, 2014. **9**(6): p. e99528.

167. Al-Attar, A., et al., *Human apurinic/apyrimidinic endonuclease (APE1) is a prognostic factor in ovarian, gastro-oesophageal and pancreatico-biliary cancers.* Br J Cancer, 2010. **102**(4): p. 704-9.

168. Jinks-Robertson, S. and H.L. Klein, *Ribonucleotides in DNA: hidden in plain sight.* Nat Struct Mol Biol, 2015. **22**(3): p. 176-8.

169. Keszthelyi, A., et al., *Mapping ribonucleotides in genomic DNA and exploring replication dynamics by polymerase usage sequencing (Pu-seq).* Nat Protoc, 2015. **10**(11): p. 1786-801.

170. Kunkel, T.A. and D.A. Erie, *DNA mismatch repair.* Annu Rev Biochem, 2005. **74**: p. 681-710.

171. Nick McElhinny, S.A., et al., *Genome instability due to ribonucleotide incorporation into DNA.* Nat Chem Biol, 2010. **6**(10): p. 774-81.

172. Gombolay, A.L., F.O. Vannberg, and F. Storici, *Ribose-Map: a bioinformatics toolkit to map ribonucleotides embedded in genomic DNA.* Nucleic Acids Res, 2018.

173. Wanrooij, P.H., et al., *Ribonucleotides incorporated by the yeast mitochondrial DNA polymerase are not repaired.* Proc Natl Acad Sci U S A, 2017. **114**(47): p. 12466-12471.

174. Berglund, A.K., et al., *Nucleotide pools dictate the identity and frequency of ribonucleotide incorporation in mitochondrial DNA.* PLoS Genet, 2017. **13**(2): p. e1006628.

175. Stuckey, S., K. Mukherjee, and F. Storici, *In vivo site-specific mutagenesis and gene collage using the delitto perfetto system in yeast Saccharomyces cerevisiae.* Methods Mol Biol, 2011. **745**: p. 173-91.

176. Sambrook, J. and D.W. Russell, *Isolation of DNA fragments from polyacrylamide gels by the crush and soak method.* CSH Protoc, 2006. **2006**(1).

177. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing reads.* EMBnet.journal, 2011. **17**: p. 10-12.

178. de Zamaroczy, M. and G. Bernardi, *The GC clusters of the mitochondrial genome of yeast and their evolutionary origin.* Gene, 1986. **41**(1): p. 1-22.

179. Dujon, B., et al., *Genome evolution in yeasts.* Nature, 2004. **430**: p. 35.

180. Wolfe, K.H., *Comparative genomics and genome evolution in yeasts.* Philos Trans R Soc Lond B Biol Sci, 2006. **361**(1467): p. 403-12.

181. Gustafsson, C.M., M. Falkenberg, and N.G. Larsson, *Maintenance and Expression of Mammalian Mitochondrial DNA.* Annu Rev Biochem, 2016. **85**: p. 133-60.

182. de Koning, A.P., et al., *Repetitive elements may comprise over two-thirds of the human genome.* PLoS Genet, 2011. **7**(12): p. e1002384.

183. Storici, F., Lewis, L.K., Resnick, M.A, *In vivo site-directed mutagenesis using oligonucleotides* Nature Biotechnology, 2001. **19**(8): p. 773-776.

184. Keskin, H. and F. Storici, *Defects in RNase H2 Stimulate DNA Break Repair by RNA Reverse Transcribed into cDNA.* Microrna, 2015. **4**(2): p. 109-16.

185. Garfinkel, D.J., et al., *Ty1 copy number dynamics in Saccharomyces.* Genetics, 2005. **169**(4): p. 1845-57.